# 論文 / 著書情報
# Article / Book Information

| | |
|---|---|
| Title | S-parameter-based modal decomposition of multiconductor transmission lines and its application to de-embedding |
| Author | S. Amakawa, K. Yamanaga, H. Ito, T. Sato, N. Ishihara, K. Masu |
| Journal/Book name | International Conference on Microelectronic Test Structures (ICMTS), , , pp. 177-180 |
| 発行日 / Issue date | 2009, 4 |
| 権利情報 / Copyright | (c)2009 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. |

# ICMTS 2009



Proceedings of the

# 2009 IEEE INTERNATIONAL CONFERENCE ON MICROELECTRONIC TEST STRUCTURES

March 30-April 2, 2009
Embassy Suites Mandalay Beach Resort
Oxnard, CA

Sponsored by the IEEE Electron Devices Society

# TABLE OF CONTENTS

**SESSION 3: Matching**
Co-chairs: Hans Tuinhout, NXP
and Mark Poulter, National Semiconductor Corp.



**SESSION 4: Poster Presentations**
**2:50 p.m. - 3:30 p.m.**
Co-chairs: Willie Sansen, KU Leuven
and Richard Allen, NIST

# High precision on-wafer backend capacitor mismatch measurements using a benchtop semiconductor characterization system

Hans Tuinhout, Fleur van Rossem[f] and Nicole Wils.

NXP semiconductors

High Tech Campus 37, 5656 AE Eindhoven, the Netherlands, hans.tuinhout@nxp.com

## Abstract

This paper discusses a sophisticated backend capacitor mismatch characterization technique based on direct capacitance measurements with a standard C-V meter, wafer prober subsite moves to measure the two capacitors of each pair sequentially and monitor the measurement noise, and statistics to take this noise appropriately into account. We describe requirements, capabilities and limitations of this approach. It is concluded that this technique proves excellently suited for assessing the matching performance of backend capacitors in the most relevant range of 10 fF to 10 pF.

## Introduction

Backend (fringe) capacitors are relatively novel integrated passive components, based on the vertical and/or lateral capacitance between stacked metal plates and/or adjacent lines (combs). The introduction of 6 to 10 layers of metallization in contemporary technology nodes, combined with aggressive shrinking of lateral dimensions make backend capacitors relatively efficient in terms of silicon area consumption (0.1 to 10 fF/$\mu m^2$). This, coupled to small voltage dependences, makes backend capacitors suitable for a wide range of analog and mixed-signal applications.

High precision mixed-signal applications such as high resolution analog-to-digital converters and filters are often based on the availability of accurately defined ratios of (switched) capacitors. Such ratios are usually realized by building arrays of unit cell capacitors [1,2,3,4], typically spanning a range of the order of 10 fF (1 unit cell) up to 10 pF (e.g. 1000 parallel unit cells). Good identicalness (matching) of these unit cells is indispensable for such applications.

For closely spaced integrated circuit devices, the matching performance is generally described in terms of the so-called Pelgrom-factor $A_{\Delta P/P}$, representing the relation between the relative mismatch fluctuation standard deviation $\sigma_{\Delta P/P}$ of parameter P and the square-root of the effective area in %$\mu$m [5]. The (stacked layers!) backend capacitor mismatch fluctuation standard deviation $\sigma_{\Delta C/C}$, is more suitably normalized relative to the population's capacitance value median, as the effective area in square microns is not the most characteristic layout quantity for a stack of metal layers. Consequently, capacitor mismatch area scaling factors ($A_{\Delta C/C}$) are generally expressed in terms of %$\sqrt{}$(fF). A typical benchmark number for $A_{\Delta C/C}$ is 1%$\sqrt{}$(fF).

For analog signal processing applications, parametric mismatch fluctuation standard deviations are required to be of the order of $10^2$ to $10^4$ ppm (0.01 to 1 %) to meet linearity specifications. At 1%$\sqrt{}$(fF), the $\sigma_{\Delta C/C}$ for a 10 fF capacitor pair should be expected to be around 0.3 % (or 30 aF), while for a population of 10 pF pairs the $\sigma_{\Delta C/C}$ could very well be 0.01 % or better. Characterization of such small capacitance mismatches places heavy demands on the Short Term Repeatability (STR) or measurement noise, to assure that mismatches are measured with sufficient precision.

Capacitor mismatch values can either be derived (indirectly) by circuit or building block designers through performance evaluation of the electronic systems that are based on them (e.g. the INL in ADCs or frequency fluctuations of loaded ring-oscillators), or they can be measured up front during process development in a parametric test environment. In the latter category, three well-known techniques are available for capacitor mismatch characterization, namely

- Direct capacitance measurements using a standard ABB (auto-balance bridge) LCR meter.
- CBCM; (charge based) capacitor load and unload currents measurements with non-overlapping clocked CMOS inverters [6]
- FGCM; floating gate capacitor matching measurement using an integrated PMOS source follower to sense the midpoint of a capacitive DC voltage divider [7].

Of these three, the first is generally considered too insensitive and/or noisy to characterize the matching of sub 100 fF capacitors with sufficient precision. CBCM was reported by the author of [6] to be prone to accuracy limitations associated with mismatch of the inverter's transistors, while the FGCM method struggles with plasma processing induced charges and damage and/or gate leakage of the source follower transistor [8,9].

**This paper** challenges the first supposition, namely that capacitor measurements with a standard C-V meter are too noisy and hence not suitable for high precision capacitor mismatch characterization. The outcome of this work is a new measurement algorithm that is based on multiple measurement repetitions and individual device measurements through prober subsite stepping, which proves this assumption wrong. In the following

[f] Fleur van Rossem is from the University of Twente, the Netherlands. She contributed to this work during an MSc internship assignment at NXP research.

sections we summarize test structure requirements, the measurement algorithm, some outstanding results and a discussion on limitations and possible pitfalls. We conclude that this technique proves excellently suited for assessing the matching performance of backend capacitors in the most relevant range from 10 pF down to as low as 10 fF.

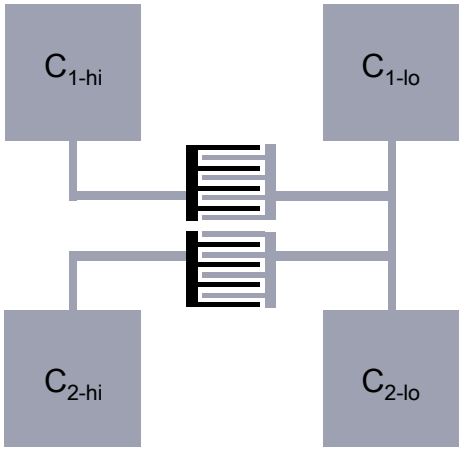



*Figure 1. Schematic representation of capacitor matched pair test structure suited for DUT1-2-1-2 characterization. The low and high sides of the capacitors in this example are drawn schematically as intertwined combs. Note however, that there are many alternatives to create backend capacitors.*

## Test Structure Requirements

The characterization technique described in this paper is based on keeping the measurement system (and hence the test structures) as simple and consistent as possible. Backend capacitor matched pair test structures should be laid out according to the following guidelines:

1. Place the two identically designed capacitors as symmetrical as possible between the two sets of probe pads (Figure 1).
2. The connection frames between the DUTs (Device Under Test) and their pads should be as symmetrical and as consistent as possible.
3. The size of high precision matched backend capacitors should be adjusted by placing multiple minimum capacitor unit cells in parallel in a cross coupled and common centroid layout. Make sure that connections, crossings and via's do not disturb the symmetry of the layout.
4. Maintain the metal density symmetry in and around arrays of matched capacitor cells. Use grounded dummy cells around the DUTs when possible.
5. Assure that the dummy tiling (metal density homogeneity improvement to facilitate chemical mechanical polishing of backend layers) is symmetrical on AND near matched capacitors. Use the 'no-tile' mask and/or manually placed metal dummy features to adhere to metal density design rules while maintaining test structure layout symmetry. Automatic chip-finishing tiling placement tools can and will result in undesirable (uncontrollable) asymmetries. This was shown to seriously affect matching of MOS transistors [10], but it needs no arguing that uncontrolled placement of floating metal tiles is bound to have even more devastating effects on the matching of high precision backend capacitors (we are looking for mismatches as low as 0.01 % !).
6. Probe pads and connection frames should be laid out only in the upper (two) metal layers to reduce parasitic capacitances.
7. Add a test module with just the pads and the frame for optimal zero-ing (open compensation) of the C-meter.

The matching measurements discussed in this paper were obtained from two different backend capacitor types fabricated in two different foundry technologies. The measurement method development and measurement noise impact assessments were mainly done using (MiM) parallel plate capacitors from a 0.18 µm technology, while the proof of the pudding results are based on backend fringe capacitor matched pairs realized in a 45 nm CMOS technology.

## Measurement Method

The test structures are measured with a Keithley 4200-SCS benchtop semiconductor characterization system with the Model 4200-CVU Integrated C-V Option and probed with 200 and 300 mm atto-guarded Cascade-Microtech wafer probers. The C-V option provides easy tailoring of DC and AC measurement conditions and functions, through the standard interactive user interface. DC Coaxial DCP100 Kelvin probes, more than adequate for C-V measurements up to 10 MHz, are used for all experiments. The C-V meter's high and low 'Cur and Pot' cables are directly connected to the SSMC 50 (Kelvin) connectors on the probe holders. For small signal capacitance measurements, the choice of the AC source signal frequency and level as well as the meter settings (measurement speed / filtering) should be chosen to fit DUT and system. Obviously a higher AC frequency and a larger AC signal level make it easier to measure (small) capacitance values. Backend capacitors are practically bias voltage independent and have relatively low resistances in the plates (fingers), allowing large AC signals and fairly high frequencies for the mismatch measurements. Unless mentioned explicitly, the AC signals were set at 99 mV at 1 MHz in all examples discussed in this paper.

Data analysis is done with MS Excel spreadsheets. Macros collect the measurement data from the hundreds of (Excel compatible) Keithley data files (one for each capacitor on each sub-site position), after which median selection, parametric position dependence evaluation and robust statistical analysis are executed and displayed in graphs.

The crux of the high precision capacitor mismatch measurement method presented in this paper is based on the following three key elements:

**1. Multiple observations**. To achieve optimum measurement performance with the highest precision, the C-V meter is set to 'quiet'. Under these conditions, the meter yields excellent low measurement noise at an average measurement speed of 0.24 s

per measurement point (Figure 2).

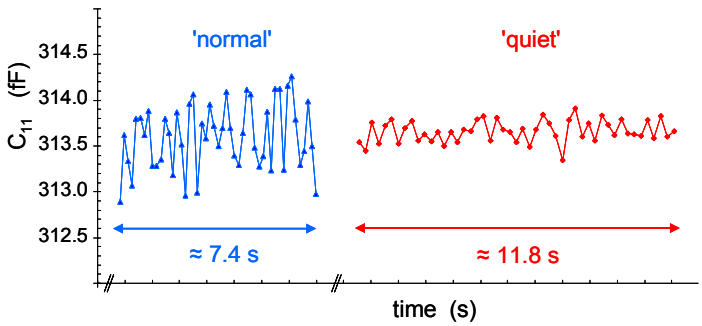



*Figure 2. Example of measurement noise of 49 capacitance observations for two meter settings for a 313.6 fF capacitor. Normal: 0.12 s/obs & σ=0.15 %; Quiet: 0.24 s/obs & σ=0.04 %.*

However, as indicated in the introduction, the expectedcapacitor mismatch standard deviations are in the range of 0.01 to 0.3 %, so this performance may not be sufficient for high precision capacitor mismatch studies. From elementary undergrad measurement classes we know that to collect more precise measurement data than those obtainable under standard meter settings, multiple observations can be used. In a white noise dominated system, the uncertainty of the C-measurement is expected to reduce with $1/\sqrt{N_{obs}}$, where $N_{obs}$ is the number of observations used for a single measurement.

The capacitor mismatch measurement algorithm used in this study was built up under the assumption that each capacitance measurement would be repeated anywhere between 9 and 999 times using the sampling mode of the measurement system. Of these multiple observations the median value (the middle value of the low-to-high sorted observations) is then used as the best guess for the capacitance.

This approach was shown before for other types of high precision mismatch measurements [7,11]. Previous studies have not shied away from increasing the number of observations up to several hundreds of measurements to reach sub 100 ppm STR levels. As this obviously results in time consuming measurement sessions for large populations of matched pairs, one should always try to limit $N_{obs}$ to such a number that the 'real' parametric mismatch fluctuation is limited by the statistical uncertainty associated with the population size, rather than by the measurement noise. As a rule of thumb, one can assume that this is generally the case when the measurement noise is less than one third of the real parametric mismatch fluctuations when the population size is of the order of 60 to 100 samples (typically the number of test dies available on one full wafer).

**2. Sub-site moves.** The two capacitors of each pair are measured sequentially by stepping from the first to the second device after the ($N_{obs}$) measurements on the first capacitor are completed. This is done with the wafer prober through a sub-site move (Figure 3). The main advantage of this is that both capacitors of the pair are measured in exactly the same way, with the same probes, cables and meter (offset).

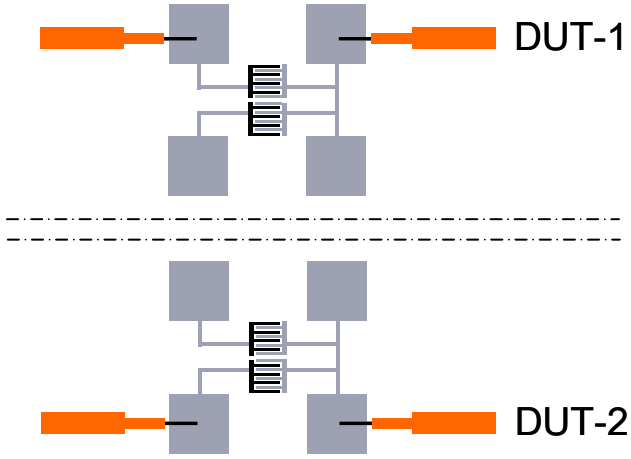



*Figure 3. The principle of DUT-1-2 prober sub-site moves for matched pair measurements: both capacitors of the pair are measured with exactly the same measurement set-up.*

**3. DUT-1-2-1-2 measurement noise correction.** The main novelty of the currently described algorithm is that each pair is measured **twice** to quantify the impact of the measurements noise. This is done by stepping backwards and forward again and re-measuring both capacitors. This so-called DUT-1-2-1-2 method allows separating the measurement noise from the true device mismatch. This is done as follows:

The relative capacitance mismatch for the **first** measurement of a particular pair i of the population is defined as

$$\Delta C/C_{1i} = 2x\ (C_{11i} - C_{21i}) / (C_{11i} + C_{21i}),$$

where $C_{11i}$ and $C_{21i}$ represent the **first** time capacitors 1 and 2 are measured for pair i. The total population of matched pair measurements then yields a relative capacitor mismatch fluctuation standard deviation estimator $\sigma_{\Delta C/C1}$. Likewise, with $C_{12i}$ and $C_{22i}$ representing the **second** measurement for each capacitor, one obtains:

$$\Delta C/C_{2i} = 2x\ (C_{12i} - C_{22i}) / (C_{12i} + C_{22i}),$$

with $\sigma_{\Delta C/C2}$ as estimator for the standard deviation of the matching for the set of second measurements. The impact of the measurement noise on the mismatch estimation for pair i is now sampled in the STR of the mismatch measurements

$$\Delta\Delta C/C_i = \Delta C/C_{1i} - \Delta C/C_{2i}$$

giving $\sigma_{\Delta\Delta C/C}$ as an estimator for its standard deviation. The essence of this is that although only one measurement noise (STR) observation is sampled for each pair, the collection STR observations resulting from the entire population (60 to 100 pairs) of these DUT-1-2-1-2 double mismatch measurements allows very reliable estimation of the measurement system noise.

If the differences between the two subsequent mismatch measurements for each pair are statistically independent from each other as well as from their actual mismatch value (which is

more than likely since we are looking at measurement noise), the short-term repeatability standard deviation for a particular measurement can be estimated as

$$\sigma_{STR} = \sigma_{\Delta\Delta C/C} / \sqrt{2}$$

and the variance of the STR can be subtracted from the measured mismatch variance to obtain a noise corrected estimator for the mismatch standard deviation $\sigma_{\Delta C/C\_cori}$ through

$$\sigma^2_{\Delta C/C\_cor1} = \sigma^2_{\Delta C/C1} - \sigma^2_{STR} \quad \text{or} \quad \sigma^2_{\Delta C/C\_cor2} = \sigma^2_{\Delta C/C2} - \sigma^2_{STR}$$

## Results

**1. Measurement algorithm characterization.** The performance of the noise corrected DUT-1-2-1-2 technique as described above is first demonstrated by an example based on two high density MiM parallel plate capacitor matched pairs built in an 0.18 µm analog & mixed signal process. The populations consist of 98 pairs spread out across a 200 mm wafer. Pair 1 is a 1.24 pF (40x40 µm$^2$) pair and pair 2 a 313 fF (20x20 µm$^2$) pair. Note that the second pair's capacitors are a factor four smaller in area and capacitance. These pairs were measured with different numbers of $N_{obs}$'s ranging from 1 to 195, and two AC signal levels. Some of the results are summarized in figures 4, 5 and 6.

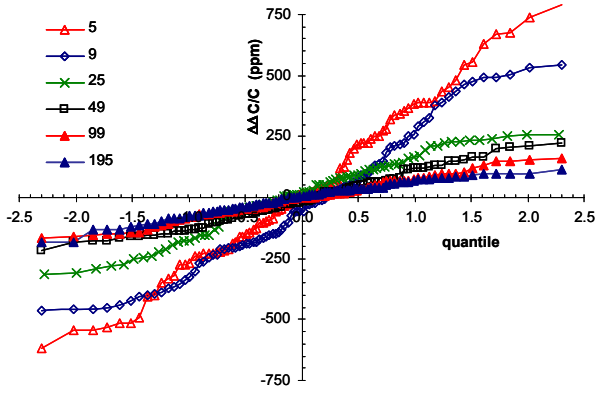



*Figure 4. Example of normal scaled cumulative probability plot of ΔΔC/C for different $N_{obs}$, varying between 5 and 195. Pair 2 (313 fF); AC signal: 99 mV @ 1 MHz.*

Figure 4 shows an initial sanity check for the method, namely whether the STR is normally distributed. The figure shows normal scaled cumulative probability plots of the mismatch short term repeatability $\Delta\Delta C/C_i$ for different numbers of $N_{obs}$ for the 313 fF pair 2. Straight lines confirm normal distributions and the slopes represent the standard deviations. As expected, $\sigma_{\Delta\Delta C/C}$ reduces with $N_{obs}$.

Figure 5 summarizes the scaling of $\sigma_{\Delta\Delta C/C}$ with $N_{obs}$ for both pairs. The repeatability indeed scales with $N_{obs}$, albeit that the expected $1/\sqrt{N_{obs}}$ behavior (dashed lines) is not fully met. It is important to note however, that the four times larger capacitors of pair 1, measured with a proportionally lower signal level yields a comparable STR at the same $N_{obs}$ as the smaller pair with the full 99 mV signal. This implies that one has two independent knobs ($N_{obs}$ and AC level) to adjust the STR to reach the required precision for a capacitor mismatch measurement. Note that for pair 2 one would need at least 200 observations to bring the STR down to a level of 100 ppm that would seem appropriate according to the rule of thumb suggested in the previous section.

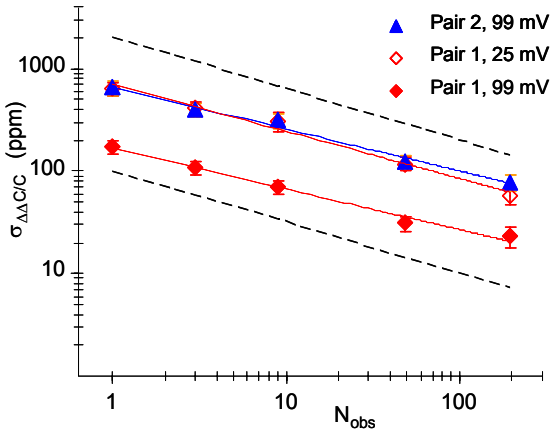



*Figure 5. STR of capacitance mismatch measurements vs. $N_{obs}$ for different capacitor sizes and AC signal levels*

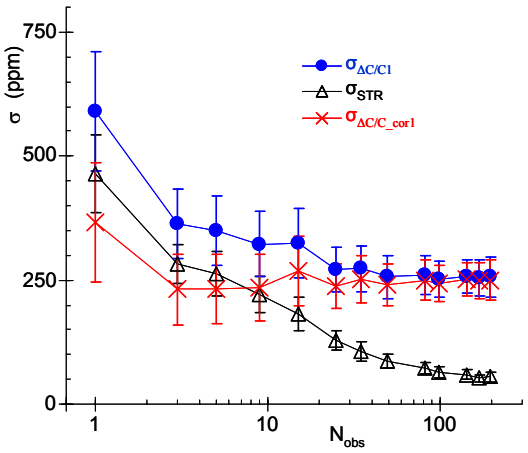



*Figure 6. $\sigma_{\Delta C/C1}$, $\sigma_{STR}$ and $\sigma_{\Delta C/C\_cori}$ vs $N_{obs}$ for pair 2 (313 fF); AC signal: 1 MHz, 99 mV; error bars represent 3σ statistical uncertainty from bootstrap analyses.*

The real strength of the noise corrected DUT-1-2-1-2 mismatch measurement method is depicted in figure 6, where $\sigma_{\Delta C/C1}$, $\sigma_{STR}$ and $\sigma_{\Delta C/C\_Cor1}$ are all plotted versus $N_{obss}$. For this population, the 'real' mismatch standard deviation estimator ($\sigma_{\Delta C/C\_Cor1}$) is apparently about 250 ppm. The 3σ uncertainty bars in figure 6 were obtained from bootstrap statistical uncertainty analyses on the measured populations of $\Delta C/C_1$'s and $\Delta\Delta C/C$'s. It is clear from this figure that a substantial number of observations is required before $\sigma_{\Delta C/C1}$ reaches its saturation level. The most striking conclusion from this experiment is however, that although one would expect that at least 50 to 100 observations would be required to obtain sufficient STR (according to the reasoning given above), already by as few as 25 observations, the mismatch fluctuation estimator is well within the statistical

uncertainty that is calculated for this population. For the brave even 3 or 9 observations would probably suffice. This means that statistics and subsequent noise correction really helps to estimate the true mismatch fluctuation standard deviation, even when the measurement noise is substantial.

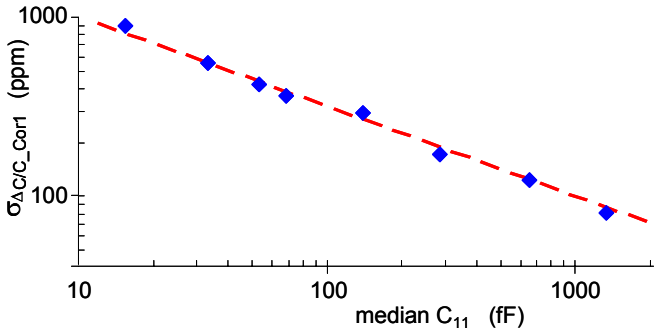



*Figure 7. Pelgrom plot of $\sigma_{\Delta C/C\_cor1}$ vs. the medians of $C_{11}$ for a collections of populations of cross coupled unit cell matched capacitors. Dashed line represents 0.32 %√fF area scaling.*

**2. The proof of the pudding.** The encouraging results of the noise corrected DUT-1-2-1-2 measurement method of the previous section are very convincingly enhanced by figure 7. This example of the backend capacitor relative mismatch fluctuation area scaling was measured using a set of 8 properly laid-out stacked metal comb (fringe cap) matched capacitors fabricated in a 45 nm CMOS technology. Population size on the 300 mm wafers is 98 pairs for each layout. The median C-value for this example ranges from 1.33 pF for the largest capacitor (144 unit C-cells) to a minimum of only 15 fF for a pair consisting of 2 (2x2 cross-coupled) unit cells, see table 1. Both the textbook-like mismatch fluctuation area scaling behavior, as well as the very acceptable 0.32 %√fF (dashed line in figure 7) for this type of devices, while going all the way down to sub 20 fF capacitors took away all reservations that stake holders in the field had against using direct C-V measurements for high precision capacitor mismatch measurements. This example indeed proves quite convincingly that the measurement noise corrected DUT-1-2-1-2 technique is very well suited for high precision backend capacitor mismatch characterization.

## Discussion

A perhaps counter-intuitive observation from figure 5 is that mismatch measurements for large capacitors are less affected by measurement noise than for small capacitors. This means that it is apparently 'easier' to measure the mismatch of large capacitor pairs than for small pairs, even though the matching of large capacitors is much better (100 ppm or lower). This is due to the larger signal that is measured in the AC meter (detector), which is proportional to the product of the capacitance value and the AC voltage level. While the mismatch standard deviation of smaller capacitor pairs increases with the inverse of the square root of the median capacitance value, the measurement noise increases inversely with its median (as the measured signal decreases linearly). This behavior is fundamentally different from mismatch characterization of for instance resistors or transistors, where the currents generally remain constant or even increase with smaller dimensions. In combination with an inverse square-root area increase of the mismatch standard deviations, parametric mismatch standard deviations of **small** devices are generally easier to measure.

*Table 1. Details and results of an example of noise corrected DUT-1-2-1-2 capacitor mismatch characterization.*

| **med_$C_{11}$ (fF)** | **# cells (-)** | **$\sigma_{\Delta C/C\_cor1}$ (ppm)** | **$\sigma_{STR}$ (ppm)** | **$N_{obs}$ (-)** |
|---|---|---|---|---|
| 1330 | 144 | 80 | 23 | 49 |
| 655 | 72 | 125 | 44 | 49 |
| 286 | 32 | 170 | 90 | 49 |
| 140 | 16 | 290 | 190 | 49 |
| 68 | 8 | 360 | 330 | 99 |
| 54 | 6 | 424 | 360 | 99 |
| 33 | 4 | 560 | 620 | 99 |
| 15 | 2 | 890 | 1370 | 99 |

With this in mind we can look back at figure 7 and table 1. Note that both on the 'low' precision side ($\sigma_{\Delta C/C} < 0.1$ %) as well as on the high precision side ($\sigma_{\Delta C/C} < 0.01$ %!!), there is no evidence of significant deviation from the expected physical 1/√fF behavior. This becomes even more striking when the data of table 1 are taken into consideration. Prior to this study we (kind of arbitrarily) decided that 49 observations would probably be enough for capacitors larger than 100 fF, while the number of observations was increased to 99 for the smaller ones. After analyzing the results, one cannot but conclude that statistics helps fantastically in this case! Table 1 reveals that the STRs for the smallest pairs are even significantly larger than their mismatch standard deviations. The largest pairs on the other hand come out with a fantastic low STR of 23 ppm with only 49 observations. This implies that there is a lot more room for measurement precision improvement on this side. Although such devices did not form part of the examples in this paper, it must be deemed more than likely that for capacitors larger then 10 pF, STRs below 10 ppm are attainable.

The reason why figure 5 does not follow a perfect $1/\sqrt{N_{obs}}$ trend is believed to be related to slow measurement system drift and/or 1/f noise in the system. Extremely long measurements (up to 1000 observations) have so far not yielded sub 10 ppm STRs for single digit pF capacitors due to an apparent saturation of the STR. Note that for 1000 observation the system takes about 4 minutes per capacitor or 16 minutes per pair, or over 26 hours for an entire population of 98 pairs, opening up plenty of room for temperature and equipment drifts and other disturbances.

As the essence of the noise corrected DUT-1-2-1-2 method lies in taking the measurement noise into account, it is obviously not

really necessary to **step back** to the capacitors after collecting multiple observations on each capacitor the first time. Theoretically as well as practically it will suffice to collect a certain number of observations for both capacitors (DUT-1-2) and then divide this number into two equal groups to calculate $\Delta C/C_{1i}$, $\Delta C/C_{2i}$ and $\Delta\Delta C/C_i$. The only difference between a DUT-1-2 method and the DUT-1-2-1-2 version lies in possible effects of probe-to-pad resistance variations that may affect the double stepped version more as this involves more probe-downs.

To avoid probing twice on exactly the same 'scratch' of the pad, we applied a minor (5 µm) offset in the landing positions when stepping back to DUTs 1 and 2. Due to a relatively large capacitive coupling between the probe needles and the test structure in that particular case (clumsy layout of test structure and pads, in combination with wrong orientation of the probe needles), a small systematic offset in $\Delta\Delta C/C_i$ was encountered, corresponding to a small but significant capacitive mismatch offset of as small as 13 aF due to the 5 µm different location of the second probe landing. In this context it must be mentioned that it proves important to assure that probe needles are always placed as good as possible in an identical orientation relative to both capacitors of the pair. Beware that the placement of the probe needles is such that they do not cross either of the two devices during either of the probe downs, giving rise to a different parasitic coupling to DUT1 compared to DUT2. A meager 100 aF offset is easily picked up due to direct cross-talk between DUT and probe.

The final remark on the limitations of the technique is related to the total measurement time that it consumes. It should be realized that a figure like Figure 7 is based of in total more than 216 thousand capacitance measurements. At 0.24 s per observation this corresponds to about 15 hours measurement system / prober time for 8 populations of one capacitor type on one wafer. From a device research as well as a process characterization / modeling standpoint this should be deemed quite acceptable, given the superb quality of the resulting mismatch fluctuation area scaling graph.

## Conclusions

This paper demonstrates that it is quite well possible to characterize high precision backend capacitor mismatch using a standard benchtop semiconductor characterization system. Test structure requirements to assure suitability for a prober sub-site stepping based measurement approach were summarized. The new measurement approach is based on combining multiple observations, prober sub-site stepping and measurement noise assessment during the mismatch measurements. By taking measurement noise properly into account, reliable and realistic estimators for the relative capacitor mismatch standard deviations can be determined. This was demonstrated on different types of matched backend capacitor pair implementations from different technologies. The method proves effective for small capacitor pairs down to values as low as 10 fF, albeit that one must use sufficiently large populations of matched pairs and rely on the help of statistics, as the measurement noise for such small capacitors can easily be larger than the mismatch fluctuation standard deviation. Although larger capacitance pairs (1 to 20 pF) may have very small mismatch standard deviations (well below 100 ppm), their mismatch characterization turns out to be relatively easy, due to the larger AC signal levels in the meter. A few points of attention and limitations of the technique are worth paying attention to, but it can be concluded that the noise corrected DUT-1-2-1-2 capacitor mismatch characterization technique proves excellently suited for assessing the matching performance of backend capacitors in the 10 fF to 10 pF range which is relevant for most high precision mixed signal applications.

## References


[1] J. McCreary and D.A. Sealer, *"Precision Capacitor Ratio Measurement Technique for Integrated Circuit Capacitor Arrays"*, IEEE Transactions on Instrumentation and Measurement, Vol 28, No 1, pp 11-17, 1979.

[2] J. McCreary, *"Matching Properties, and Voltage and Temperature Dependence of MOS Capacitors"*, IEEE Journal of Solid State Circuits, Vol 16, No 6, pp 608-616, 1981.

[3] R. Singh and A.B. Bhattacharyya, *"Matching Properties of Linear MOS Capacitors"*, Solid State Electronics, Vol 32, No 4, pp 299-306, 1989.

[4] M.J. McNutt, S. LeMarquis and J.L. Dunkley, *"Systematic Capacitance Matching Errors and Corrective Layout Procedures"*, IEEE Journal of Solid State Circuits, Vol 29, No 5, pp 611-616, 1994.

[5] M.J.M. Pelgrom, A.J. Duinmaijer and A.P.G. Welbers, *"Matching Properties of MOS Transistors"*, IEEE Journal of Solid State Circuits, Vol 24, No 5, pp 1433-1440, 1989.

[6] J.C. Chen, D. Sylvester, C. Hu, H. Aoki, S. Nakagawa and S-Y. Oh , *"An On-Chip, Interconnect Capacitance Characterization Method with Sub-Femto-Farad Resolution"*, ICMTS97 Proceedings of the IEEE International Conference on Microelectronic Test Structures, Vol. 10, pp. 77-80, 1997.

[7] H.P. Tuinhout, H. Elzinga, J.T. Brugman and F. Postma, "*The Floating Gate Measurement Technique for Characterization of Capacitor Matching"*, IEEE Transactions on Semiconductor Manufacturing, Vol 9, No 1, pp 2-18, 1996.

[8] O. Roux dit Buisson, G. Morin, F. Paillardet and E. Mazaleyrat, *"A New Characterization Method for Accurate Capacitor Matching Measurements using Pseudo-Floating Gate Test Structures in Submicron CMOS and BiCMOS Technologies"*, ICMTS98 Proceedings of the IEEE International Conference on Microelectronic Test Structures, Vol. 11, pp. 223-227, 1998.

[9] Weidong Tian et al, "*Gate Oxide Leakage and Floating Gate Capacitor Matching Test"*, ICMTS07 Proceedings of the IEEE International Conference on Microelectronic Test Structures, pp. 120-123, 2007.

[10] H.P. Tuinhout and M. Vertregt, *"Characterization of Systematic MOSFET Current Factor Mismatch Caused by Metal CMP Dummy Structures"*, IEEE Transactions on Semiconductor Manufacturing, Vol 14, No 4, pp 302-310, 2001.

[11] T. Ewert, H.P. Tuinhout, N.A.H. Wils and J. Olsson *"Design and implementation of an ultra high precision parametric mismatch measurement system"*, ICMTS05 Proceedings of the IEEE International Conference on Microelectronic Test Structures, pp. 149-154, 2005.

# Static Noise Margin Evaluation Method Based on Direct Polynomial-Curve-Fitting with Universal SRAM Cell Inverter TEG Measurement

Kazuyuki Nakamura, Kazunori Noda, and Hiroki Koike*

Center for Microelectronic Systems, Kyushu Institute of Technology
* Fukuoka Industry, Science & Technology Foundation
680-4 Kawazu, Iizuka, Fukuoka 820-8502, Japan
Tel: +81-948-29-7584, Fax: +81-948-29-7586, Email: nakamura@cms.kyutech.ac.jp

## ABSTRACT

A new method to evaluate the Static Noise Margin (SNM) for leading-edge CMOS SRAM development is proposed. This method includes: (1) direct measurement of the inverter DC transfer curves using a "Universal SRAM Cell Inverter TEG (USCIT)" with arbitrary transistor ratios, (2) curve-fitting of the measured data to polynomial functions in a 45-degree rotated space, and (3) a database of the polynomial coefficients to evaluate and optimize the SNM by a simple algebraic operation. The SNM values obtained using this method are in good agreement with the measured SRAM operations.

## INTRODUCTION

In the development of SRAMs for the 45-nm technology generation and beyond, it is difficult to maintain a sufficient static noise margin (SNM) in the SRAM because of supply voltage scaling and device variability [1][2]. Historically SNM analysis using classical MOSFET Id-Vd expressions has been used, but it has been shown that this cannot fit the scaled CMOS technologies. A method that uses SPICE simulation is now more widely used; the procedure for this is depicted in Fig. 1 [3][4]. This method comprises the following steps: (i) measuring I-V characteristics for a single transistor TEG, (ii) extracting the SPICE model parameters, (iii) simulating DC transfer curves of the inverter in an SRAM cell, and (iv) drawing the butterfly curves by combining the two inverter characteristics and reading off the SNM value. This method is, however, inadequate in several aspects, such as fitting errors in the SPICE model parameter extraction and the accuracy limit of the SPICE simulation. This imposes considerable restrictions on the design of highly scaled CMOS SRAM.

## NEW SNM EVALUATION METHOD

The proposed method, illustrated in Fig. 2, comprises the following steps: (1) direct measurement of inverter DC transfer curves using an USCIT that can realize arbitrary transistor ratios, (2) fitting of the measured data to polynomial functions in a 45-degree rotated space, and (3) storing the coefficients of the polynomial functions in a database to evaluate and optimize the SNM by a simple algebraic operation. Since this method is based on measured data and directly curve-fitted polynomials, neither SPICE model extraction nor other accuracy degradation steps are included. In addition, the SNM can be obtained simply by subtracting the coefficients of two polynomials which can be stored in a database based on the measured data. This method requires neither SPICE simulation nor drawing butterfly curves, therefore, it is suitable for statistically aware memory cell optimizations.

## DESIGN OF UNIVERSAL SRAM CELL INVERTER TEG (USCIT)

A typical 6-transistor SRAM cell circuit is shown in Fig. 3. To calculate the SNM, the DC transfer characteristics of the inverter with three transistors, namely Load PMOS (PL), Driver NMOS (ND), and Transfer NMOS (NT), must be evaluated. We have developed the USCIT that can evaluate arbitrary transistor ratios for these three transistors with selectable binary-weighted transistor arrays as shown in Fig. 4. The gate width (W) can be set by 5-bit signals for PLi, NDj, and NTk (where i, j, k=0~4), respectively. We design the five gate widths for the weighted transistor array: 1/1.5/2/4/8 x Wmin, instead of a simple binary-weighted array using Wmin as the base. Wmin is the minimum transistor gate width determined by process technology. By combining these widths, we can successfully achieve any transistor size covering 1.0 ~ 15.5 times the value of Wmin with a minimum resolution of half the value of Wmin. As a result, about 30,000 design combinations can be examined with this USCIT. Fig. 5 shows the layout of an USCIT using 0.18-um technology. The macro size is 17 x 108 $um^2$. The measured DC transfer characteristics for the USCIT are shown in Fig. 6. By setting the level of the BL terminal to VDD or GND, we obtain the DC transfer curves (Fig.6 (a) and (b)) corresponding to the SRAM READ/WRITE operations, respectively.

## POLYNOMIAL-CURVE-FITTING IN A 45-DEGREE ROTATED SPACE

A method that utilizes Butterworth filter function for DC transfer curve fitting was proposed [5]. However, we tried

to fit the DC transfer curves measured by the USCIT to simpler polynomial functions. Since parts of the curve are parallel to the X- or Y-axis as shown within the dotted area in Fig. 7(a), it would be difficult to fit the curve accurately even if the degree of the polynomial were increased. In order to solve this issue, we rotate the measured curve by 45 degrees in the X-Y coordinates before the polynomial curve fitting. Table 1 compares the accuracy of the curve fitting between the original and rotated curves. Rotating the curve improves the accuracy significantly. The error can be suppressed down below 1% by extending the polynomial to the degree of 4. The results of fitting a curve to a polynomial function of degree of 5 are shown in Fig. 7(b).

## SNM Calculation with Polynomial Database

The SNM of an SRAM cell is calculated from the polynomial functions of the 2 inverters that compose the SRAM cell. Here we assume that the 2 characteristics are expressed by the following polynomial functions;

$$y_1(x) = (1/\sqrt{2}) \sum_{k=0}^{n} (a_{1k} \cdot x^k) \quad \ldots (1)$$

$$y_2(x) = (1/\sqrt{2}) \sum_{k=0}^{n} (a_{2k} \cdot x^k) \quad \ldots (2)$$

The SNM function, snm(x), is defined by subtracting $y_2(-x)$ from $y_1(x)$. Using Eqs. (1) and (2), we obtain

$$\begin{aligned} snm(x) &= y_1(x) - y_2(-x) \\ &= (1/\sqrt{2}) \sum_{k=0}^{n} [\{a_{1k} - (-1)^k \cdot a_{2k}\} \cdot x^k] \quad \ldots (3) \end{aligned}$$

An example of function snm(x) is plotted in Fig. 8. The SNM is the smaller of the absolute values of the 2 extrema (maximum and minimum values) in Eq. (3). Note that if the polynomial functions are fitted into 5-degrees or less, the algebraic solution can be used to find the extrema. As a result, the corresponding SNM value can be calculated instantly by a simple algebraic operation on the polynomial coefficients in the fitted-curve database.

## Experimental Results

We have fabricated a 256-bit SRAM chip, in which the unit cell is composed of two USCITs as shown in Fig. 4. The chip layout is shown in Fig. 9. Figs. 10(a), (b), and (c) show correlations between the SNM values calculated from the conventional SPICE simulation method, SNM values obtained from the measurement results of the USCIT, and PASS/FAIL values of an entire 256-bit SRAM TEG chip operation, respectively. The horizontal and vertical axes denote the gate widths for Transfer NMOS (NT) and Load PMOS (PL), respectively. The Drive NMOS (ND) size is fixed in all cases. In Figs. 10(a) and (b) the smaller of the Write or Read SNM values [in mV] is written in each block, with Write SNM values appearing in a Italic font and Read SNM values in a solid font. Each black and hatched block denotes the region where SNM < 0 and 0 ≤ SNM < 60mV, respectively. In Fig. 10(c), 'P' and 'F' indicate Pass and Fail operations, respectively. According to these results, the measured operational margin of a 256 bit SRAM (Fig.10(c)) is in good agreement with the SNM evaluation from the USCIT measurement (Fig.10(b)), when taking into account the cell transistor mismatch (i.e., 60mV for this measurement).

## Conclusion

We have developed a new SNM evaluation method based on the measurement results of an USCIT, polynomial curve fitting with a 45-degree rotation, and a polynomial coefficients database. This method has demonstrated greater accuracy than the conventional method that uses SPICE simulation, and is suitable for data processing covering the comprehensive conditions such as transistor-sizes, voltage, temperature, etc. The proposed method is, therefore, especially effective in establishing reliable guidelines for the early-stage development of highly scaled CMOS SRAM cells.

## Acknowledgments

The VLSI chip design and fabrication were supported by the VLSI Design and Education Center (VDEC), University of Tokyo, in collaboration with Rohm Corporation and Toppan Printing Corporation. This research was supported by funds from the Japanese Ministry: MEXT via the Kyushu knowledge-based cluster project. The authors thank NEC Micro Systems, Ltd. for their useful suggestions.

## References

[1] S. Inaba H. Kawasaki, K. Okano, T. Izumida, A. Yagishita, A. Kaneko, K. Ishimaru, N. Aoki, and Y.Toyoshima, "Direct evaluation of DC characteristic variability in FinFET SRAM cell for 32 nm node and beyond," *IEDM Tech. Digest*, pp. 487 - 490, Dec. 2007.

[2] M. Miyamura, T. Fukai, T. Ikezawa, R.Ueno, K.Takeuchi, and M. Hane, "SRAM critical yield evaluation based on comprehensive physical / statistical modeling, considering anomalous non-Gaussian intrinsic transistor fluctuations," *Symposium on VLSI Technology Tech. Digest*, pp. 22 - 23, Jun. 2007.

[3] E. Seevinck, F. J. List, and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells," *IEEE J. Solid-State Circuits*, vol. SC-22, no. 5, pp. 748-754, Oct. 1987.

[4] A. J. Bhavnagarwala, X. Tang, and J. D. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," *IEEE J. Solid-State Circuits*, vol. 36, no. 4, pp. 658-665, Apr. 2001.

[5] Q. Chen, A. Guha and K. Roy, "An accurate analytical SNM modeling technique for SRAMs based on Butterworth filter function," *Proc. of the 20th International Conference on VLSI Design*, pp. 615-620, Jan. 2007.

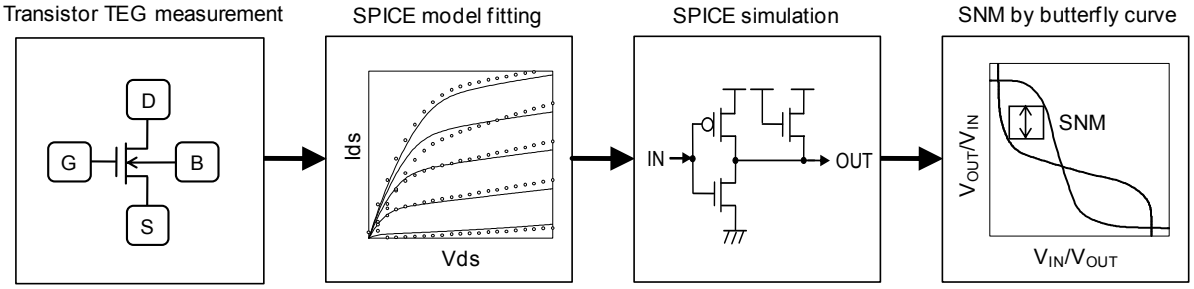



Fig. 1 Conventional method for calculating SNM values

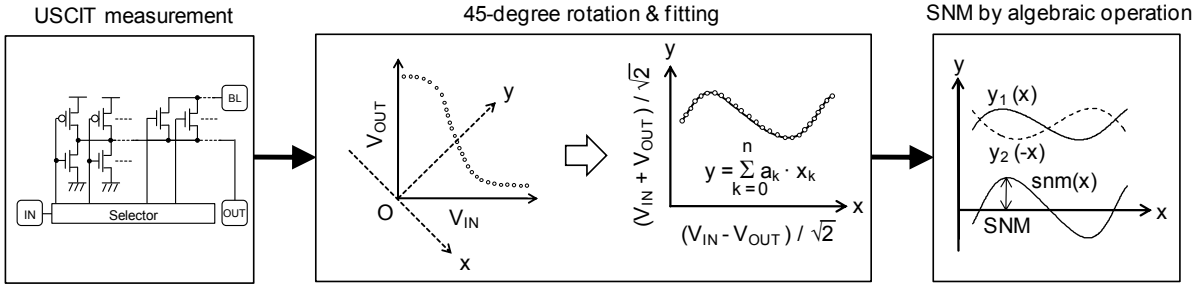



Fig. 2 The proposed method for calculating SNM values

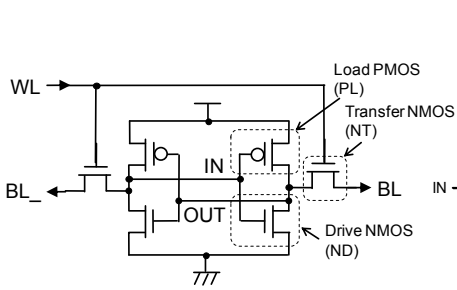



Fig. 3 6T-SRAM cell circuit

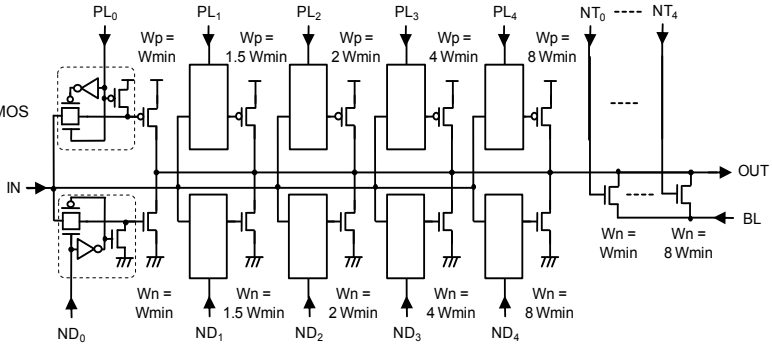



Fig. 4 USCIT circuit to measure inverter DC transfer characteristics

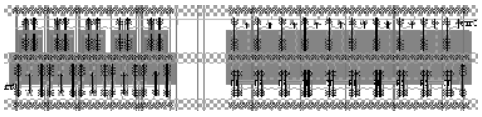
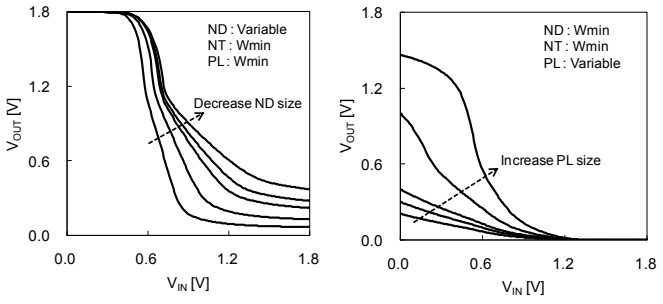


Fig. 5 Layout of the USCIT

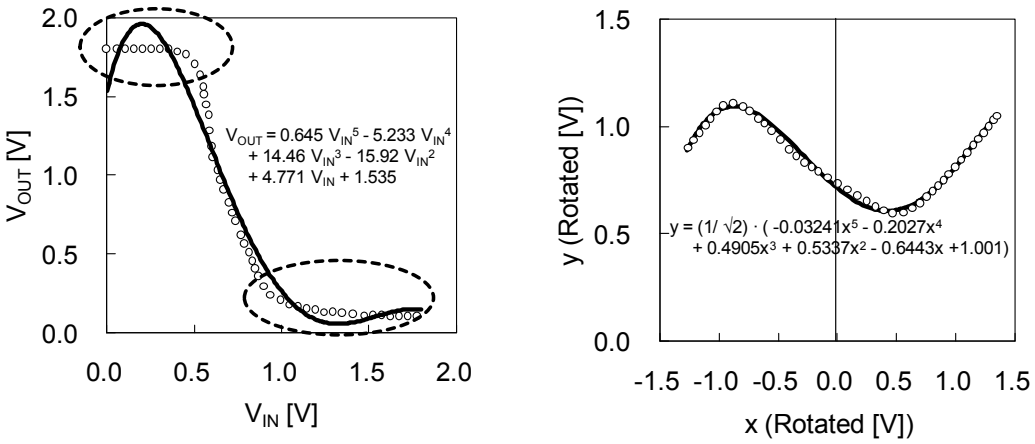
TABLE I
FITTING ERROR VERSUS THE DEGREE OF THE POLYNOMIAL USED FOR FITTING

| Degree of polynomial | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|
| Original curve fitting [%] | 10.4 | 5.7 | 5.6 | 4.0 | 2.8 | 2.8 | 2.1 |
| 45-degree rotated curve fitting [%] | 2.0 | 0.7 | 0.7 | 0.6 | 0.6 | 0.2 | 0.2 |




(a) BL=VDD (b) BL=GND

Fig. 6 Measured inverter DC transfer characteristics




(a) Original (b) After 45-degree rotation

Fig. 7 Examples of DC curve fitting to a polynomial of degree of 5
(Circles indicate measured data, while the solid line is the fitted curve.)

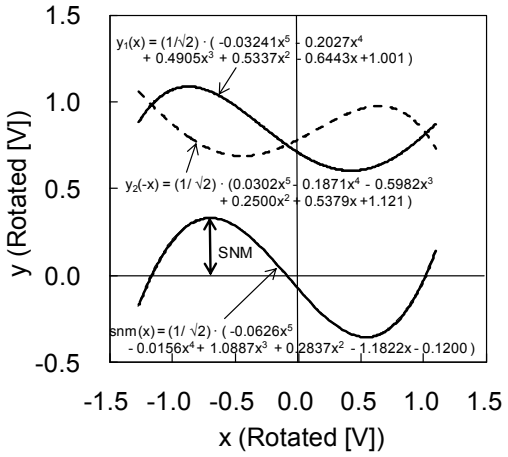



Fig. 8 An Example of SNM calculation

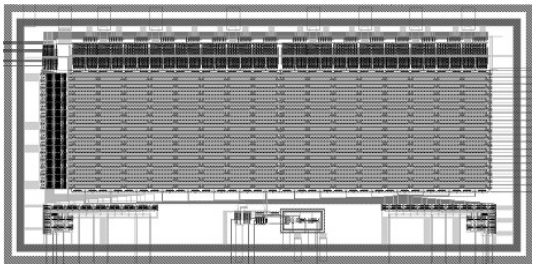


Fig. 9 256-bit SRAM TEG layout using 0.18-um technology

W of NT / W of PL

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 268 | 169 | 31 | 4 | <0 | <0 | <0 | <0 | <0 | <0 | <0 |
| 287 | 190 | 54 | 28 | <0 | <0 | <0 | <0 | <0 | <0 | <0 |
| 305 | 228 | 97 | 67 | <0 | <0 | <0 | <0 | <0 | <0 | <0 |
| 316 | 224 | 93 | 71 | <0 | <0 | <0 | <0 | <0 | <0 | <0 |
| *138* | 252 | 125 | 101 | 22 | <0 | <0 | <0 | <0 | <0 | <0 |
| *84* | 260 | 137 | 113 | 35 | 6 | <0 | <0 | <0 | <0 | <0 |
| *<0* | 268 | 160 | 138 | 62 | 33 | <0 | <0 | <0 | <0 | <0 |
| *<0* | *224* | 169 | 146 | 71 | 43 | <0 | <0 | <0 | <0 | <0 |
| *<0* | *145* | 187 | 165 | 92 | 65 | 17 | <0 | <0 | <0 | <0 |
| *<0* | *117* | 194 | 172 | 101 | 73 | 26 | 5 | <0 | <0 | <0 |

(a) SNM for Conventional method (using SPICE simulation)

W of NT / W of PL

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 320 | 231 | 114 | 101 | 28 | 1 | <0 | <0 | <0 | <0 | <0 |
| 338 | 249 | 135 | 120 | 48 | 17 | <0 | <0 | <0 | <0 | <0 |
| *283* | 283 | 170 | 157 | 86 | 57 | 12 | <0 | <0 | <0 | <0 |
| *269* | 279 | 166 | 153 | 82 | 53 | 8 | <0 | <0 | <0 | <0 |
| *123* | 304 | 194 | 180 | 111 | 82 | 39 | 20 | <0 | <0 | <0 |
| *70* | 312 | 202 | 189 | 121 | 92 | 48 | 30 | 0 | <0 | <0 |
| *<0* | *268* | 221 | 210 | 142 | 114 | 70 | 52 | 23 | 4 | <0 |
| *<0* | *218* | 230 | 218 | 151 | 123 | 80 | 59 | 33 | 19 | <0 |
| *<0* | *137* | 245 | 233 | 167 | 140 | 97 | 79 | 50 | 38 | 16 |
| *<0* | *109* | 250 | 238 | 173 | 145 | 103 | 85 | 56 | 43 | 22 |

(b) SNM for Proposed method

W of NT / W of PL

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| P | P | P | P | F | F | F | F | F | F | F |
| P | P | P | P | F | F | F | F | F | F | F |
| P | P | P | P | P | F | F | F | F | F | F |
| P | P | P | P | P | F | F | F | F | F | F |
| P | P | P | P | P | P | F | F | F | F | F |
| P | P | P | P | P | P | F | F | F | F | F |
| F | P | P | P | P | P | P | F | F | F | F |
| F | P | P | P | P | P | P | F | F | F | F |
| F | P | P | P | P | P | P | P | F | F | F |
| F | P | P | P | P | P | P | P | F | F | F |

(c) Measured PASS/FAIL with 256-bit SRAM Operation

Fig. 10. Experimental results

# Addressable Arrays Implemented with One Metal Level for MOSFET and Resistor Variability Characterization

Mark B. Ketchen, Manjul Bhushan* and Greg Costrini*

IBM Research, T. J. Watson Research Center, Route 134, Yorktown Heights NY 10598, USA
*IBM Systems and Technology Group, 2070 Route 52, Hopewell Junction, NY 12533, USA

## ABSTRACT

Addressable array test structures for rapid collection of statistical distributions of MOSFET parameters and parasitic resistances are described. A unique feature of these designs is that they require only one level of metal, yet are compact for placement in the scribe line for early process learning. MOSFET measurements are made over full range of I-V characteristics including leakage currents of individual devices in the sub-threshold region. A modular approach for test structure integration and parallel testability enables high efficiency in design and data acquisition.

## INTRODUCTION

With CMOS scaling, there is an increasing need for characterization and modeling of statistical variations in MOSFET characteristics [1]. Such variations arise from, for example, random dopant fluctuations, line-edge roughness and systematic spatial effects. It is important to model and manage such variability early in the technology development cycle as we transition from 45 nm to 32 nm technology nodes and beyond. Test structure designs comprising arrays of devices reported earlier are typically complex with several metal layers and best suited for a mature technology or occupy a large area [2-11]. Another approach, that uses compact designs with two levels of metal, does not allow accurate measurement of sub-threshold currents [12].

In this work, we describe very compact test structure designs for collecting modest statistics on the full range of MOSFET characteristics as well as resistive elements such as metal-silicon contacts and metal wires. The design of an entire test structure macro, comprising 6 to 8 arrays, each of 30 Devices Under Test (DUTs), is accomplished with only a single metal level and is suitable for placement in the scribe line for early learning in the technology development or manufacturing cycle. Macros are assembled in a modular fashion to facilitate design of experiments. Multiple arrays within a macro can be tested in parallel for rapid data collection. The design flexibility and data integrity has been demonstrated in the 45 nm and 32nm technology nodes. Both PD-SOI (Partially-Depleted Silicon on Insulator) and bulk versions of these designs are currently in use at IBM.

## DESCRIPTION OF TEST STRUCTURES AND METHODOLOGY

The basic building block of our test structures comprises an electrically one-dimensional array of 30 DUTs, circuitry for supplying appropriate currents and voltages and a 5 bit decoder to select any one, all, or none of the DUTs for measurement. This building block, or array-unit, is placed within a 60 μm x 40 μm area between two contact pads which provide connections to all 30 DUTs in parallel. For MOSFET array-units, the source and drains of all DUTs are connected to these common adjacent pads and the decoder outputs activate the voltages applied to the individual gates of the DUTs. For resistor array-units, the voltage apply/current measure leads of all DUTs are connected to the common adjacent pads and the decoder outputs activate the current through the DUT and the voltage readout across the DUT.

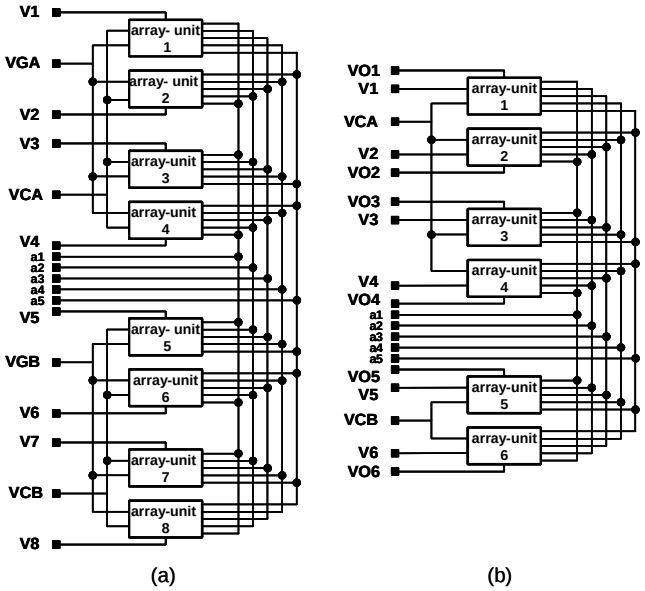



Figure 1: Top level schematic for (a) MOSFET array macro and (b) resistor macro. Decoder inputs are a1 – a5, and VCA and VCB are clamp voltages that reduce background currents from unselected DUTs. In (a) V1 – V8 are drain inputs and VGA and VGB are gate voltages. In (b) V1 –V6 are drive voltages and VO1 – VO6 are output voltages. In both macros multiple shared connections such as MOSFET source pads and decoder power supply pads are not shown.

Multiple array-units with shared decoder inputs and decoder power supplies are placed within one macro. The assembly of array-units in a macro is illustrated in Fig. 1(a) for MOSFETs and Fig. 1(b) for resistors. A MOSFET macro comprises 4 n-FET and 4 p-FET array-units while a resistor macro comprise 6 array-units. Each macro employs 25 I/O pads for testing on a standard DC parametric tester in the manufacturing line. The macro dimensions are 80 μm x 2500 μm (or 2000 μm) including the 25, 60 μm x 60 μm (or 40 μm x 40 μm) pads. The designs are multi-port, allowing test of multiple arrays within a macro in parallel. For testing on a parallel tester with 100 or more channels, the macros may be stacked or assembled to share more pads, thereby increasing the test efficiency. Many macros with a standard template are generated by populating the array-units with different DUTs defined as parameterized cells. The DUTs may differ in device type, dimensions and layout style for Design for Manufacturing (DFM) or modeling applications. A number of identical DUTs are grouped within an array or across arrays within the same macro to get random statistical variations. The variations in a parameter mean of a group of identical DUTs across wafer, wafer-to-wafer and lot-to-lot give information related to systematic process variations.

The physical layout of the MOSFET array-unit is indicated in Fig 2. The DUTs are arranged in a 3 x 10 matrix powered (drain to source) by interdigitated combs emanating from the adjacent I/O pads. The parasitic voltage drop is negligible just as for a discrete DUT placed between two pads. Vertical metal wires traverse a significant fraction of the macro height for shared decoder VDD (n-FET) and decoder GND (p-FET), decoder input bits and low current (< 1μA) DC voltage inputs to the DUTs. The decoder shares the DUT source pad, but draws only a few μA after settling from a selection transient, and does not impact DUT current measurements that are made on the drain side after the settling transient. The most challenging part of the design is the decoder which through subsequent circuitry sets gate voltages of all DUTs in the array. Wire cross-overs and much of the data-flow wiring is done in horizontal sections of the high resistance (silicided) polysilicon (poly) gate layer, while vertical metal wire segments are placed within split VDD and GND busses or adjacent to these busses. The decoder logic is implemented with inverters and two-way NAND gates comprised of single finger low-leakage devices. The gate poly is unidirectional and on a fixed pitch for both 45 nm and 32 nm technologies. In the case of 32 nm technology the metal-to-Si contacts are also gridded. The physical size of this highly customized decoder sets an upper limit of 5 decode input bits and 30 DUTs per array-unit. The remaining 2 of 32 decoder output signals are used to activate all DUTs in parallel under different bias conditions, facilitating background current correction.

In the MOSFET array-unit, the outputs of the decoder feed directly into low leakage bias circuits that steer the appropriate voltages to the DUT gates [8], as shown in Figs. 3 for n-FET DUTs. The layout depicted in Fig. 3(b) exemplifies how the wiring is accomplished with a combination of one level of metal and silicided gate polysilicon. A voltage VG is applied to the gate of a selected n-FET while the gates of the other 29 n-FETs are biased at a clamp voltage VC. By setting VC at a small negative voltage, the leakage current of the unselected MOSFETs is greatly reduced, although this reduction is ultimately limited by Gate-Induced Drain Leakage, GIDL [13]. The drain-to-source leakage currents through the bias circuits are minimized by using groups of three pass-transistors in series rather than single pass-transistors. These currents are of some concern because they ultimately flow through the VG and VC supply lines and any voltage drop in the VG line would be directly reflected as an error in the VG applied to the DUTs. The same VG bus services four array-units and utilizes resistive poly underpasses to cross the VD busses evident in Fig. 2. By keeping the leakage currents low and taking care with the VG bus design, the VG bus voltage drop is kept below 1mV. The bias circuits shown in Fig. 3 are for n-FET DUTs where the negative clamp voltage VC is typically set at -0.1 to -0.2V. In the case of p-FET DUTs the

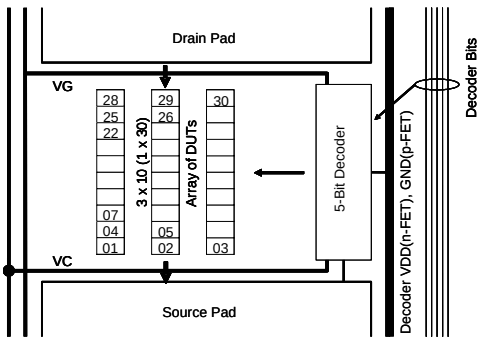



Figure 2: Physical representation of an n-FET array-unit showing the DUT locations in the 3x10 matrix, the decoder, the nearby source and drain pads, and associated I/O wiring.

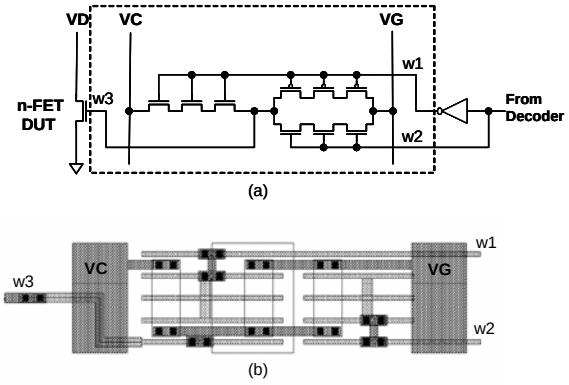



Figure 3: Low-leakage n-FET gate bias circuit. Voltage VG is applied to the selected gate and VC to the unselected gates.

bias circuits are configured differently (p-FET pass-transistors substituted for n-FET pass-transistors) to allow for a positive VC, which is typically set at 0.1 to 0.2V.

With an appropriate value of VC applied, the error in a drain-source current, $I_{ds}$, measurement due to the background current from the unselected MOSFETs is negligible when the selected MOSFET is in the on-state. However, to get the selected MOSFET's $I_{ds}$ in the sub-threshold region, a correction must be applied to the measured current. Two additional measurements are made to facilitate this correction. The $I_{ds}$ for all DUTs is measured in parallel with an applied gate-drain voltages of VC, $I_{all\_VC}$, and with an applied gate-drain voltage of VG, $I_{all\_VG}$ for all MOSFETs. For nominally identical DUTs, this correction, which is subtracted from the measured current, is 29/30* $I_{all\_VC}$. For example, assuming the ratio $I_{k\_VC}/I_{all\_VC}$ is < 0.1, where $I_{k\_VC}$ is the drain-source current of a selected DUT k with its gate at VC, this approach should give individual DUT current values with a maximum error of about 5% of the true values. The situation is more complicated if the DUTs are by design significantly different from each other. Let the true drain-source current flowing through selected DUT k with gate voltage VG be $I_{k\_VG}$, expressed as

$$I_{k\_VG} = (I_{k\_VG} - I_{k\_VC}) / (1 - \eta_k) \quad (1)$$

where $\eta_k = I_{k\_VC}/I_{k\_VG}$.

Because of the background current of the 29 unselected MOSFETs, neither $I_{k\_VC}$ nor $I_{k\_VG}$ can be directly measured. However, the difference $(I_{k\_VG} - I_{k\_VC})$ is identically equal to the measured current difference $(I_{k\_VG\_VC} - I_{all\_VC})$ where $I_{k\_VG\_VC}$ is the measured $I_{ds}$ of DUT k at VG, with the gates of all other 29 DUTs biased at VC. Furthermore, if all DUTs have identical I-V curves or more generally have similar slope in the sub-threshold region, then $\eta_k = \eta_{all} = I_{all\_VC}/I_{all\_VG}$. Under these conditions it then follows that $I_{k\_VG}$ can be determined as

$$I_{k\_VG} = (I_{k\_VG\_VC} - I_{all\_VC}) / (1 - \eta_{all}) \quad (2)$$

for all DUTs over the entire sub-threshold region, where the value of $\eta_{all}$ must be determined for each value of VG.

For resistance measurements, the array-unit is configured similar to the MOSFET array-unit with the DUTs arranged in a 3x10 matrix. Additional circuitry, shown in Fig. 4, for resistance measurements is also placed between the interdigitated drive (VDr) and GND fingers. The GND contact for all resistor DUTs is common and current is supplied via the VDr bus only to the decoder selected DUT through two series pass-transistors (NS1). Unselected DUTs are each isolated by their NS1 pass-transistors with gates at a negative clamp voltage, VC. The voltage VO across the selected DUT is read through two additional pass-transistors (NS2) with the unselected DUTs isolated from the common VO output lead by their NS2 pass-transistors with gates also set at VC. In this three terminal measurement, a correction for the GND lead resistance can be determined by shorting DUT locations #1 and #30 to GND and assuming that this resistance increases linearly with DUT number.

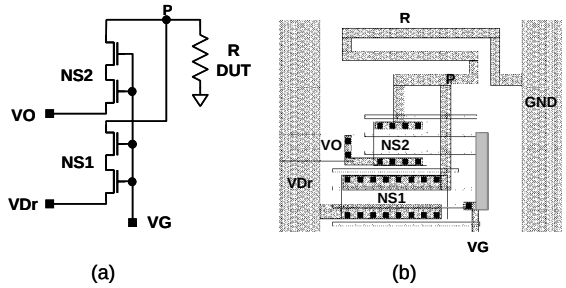



Figure 4: (a) Circuit schematic (b) and physical layout showing resistor measurement scheme for a selected DUT consisting of a segment of wire between point P and GND.

## EXPERIMENTAL RESULTS AND DISCUSSION

These array test structures are currently in use at IBM with 45 nm and 32 nm bulk and PD-SOI technologies. The validity of the macro designs was verified by simulations with full parasitic extraction. DUT sizes were selected to keep the maximum drain current to be measured < 1 mA. For 30 identical DUTs in an array-unit, the maximum simulated spread for an $I_{ds}$ value of 1 mA was 0.3 %. In the base macro design, arrays were populated with a variety of DUTs to check for any systematic variations across the array. Multiple copies of the macros were designed to cover different DUT types and configurations for DFM applications. Representative data are presented for experimental hardware in the 45 nm PD-SOI technology node to demonstrate measurement integrity.

All measurements reported here were made on IBM's standard in-line Agilent 4073A parametric tester. For MOSFET measurements, VC was typically set at -0.2V for the n-FETs and +0.2V for the p-FETs. For a given drain voltage, the threshold voltage, $V_t$ was determined as the gate voltage to achieve a fixed drain current equal to $I_{sv}$* W/L where W and and L are MOSFET widths and channel-length respectively and $I_{sv}$ is in the range of 20 nA to 300 nA. By measuring $V_t$ at two different values of drain currents, the sub-threshold slope, S, was calculated as

$$S = (V_{t1} - V_{t2}) / \mathrm{Log}_{10}(I_{t1}/I_{t2}) \quad (3)$$

where $I_{t1}$ and $I_{t2}$ are the two drain currents settings for measuring $V_{t1}$ and $V_{t2}$ [13]. Currents in the sub-threshold region and $V_t$ were also measured for all MOSFETs in each array in parallel. Note that the $V_t$ of all MOSFETs in parallel

is the geometric mean of the $V_t$ values of individual MOSFETs and is therefore smaller than the arithmetic mean of the individual Vt values, a relationship that is routinely checked to verify self consistency.

Figure 5 shows the mean $V_t$ of 64 chips on a wafer in the array A1, populated with three different $V_t$ type n-FET DUTs by column, and the second array A2, with all DUTs having the same $V_t$ type. The small variation in the mean $V_t$ with DUT location for nominally identical DUTs is within the expected range. The mean and standard deviation of $V_t$'s of the MOSFETs in an array also correlated well with the measurements made on single isolated MOSFETs of the same design on the same wafer. For $V_t$ measurements, |VC| = 0.2V and no background correction is necessary.

In Fig. 6 measured and corrected drain-to-source leakage current $I_{off}$ ($I_{ds}$ for VG=0) for DUTs in the array A1, with the three different $V_t$ types, corresponding to Fig. 5 are shown. With |VC| = 0.2V, the value of $\eta_{all}$ here and in 45 nm and 32 nm PD-SOI technology nodes in general is < 0.02 and GIDL is small, so the scheme for background correction using Eq. (2) works well. A self consistency check can also be done by comparing the measured $I_{off}$ for all DUTs in an array in parallel with the sum of the corrected $I_{off}$ values of individual DUTs. For the hardware reported on here, a variation in sub-threshold slope, S, was observed as illustrated in Fig. 7. Table 1 shows the calculated sensitivity of $\eta_k$ to changes in sub-threshold slope for |VC| = 0.1V and 0.2V with VG = 0V. Assuming 100 mV/decade to be the average slope, Table 1 also tabulates the error in the estimate of $I_{off}$ ,using Eq. (2), resulting from this variation in sub-threshold slope. With $\eta_{all}$ = 0.02 the error in $I_{off}$ is < 2% from such sub-threshold slope variations.

Table 1
Values of $\eta_k$(|VC|) with VG = 0 for different values of S in mV/decade (mV/d) and corresponding % error E(|VC|) in corrected $I_{off}$ .

| S | $\eta_k$(0.2V) | E(0.2V) | $\eta_k$(0.1V) | E(0.1V) |
|---|---|---|---|---|
| 80mV/d | 0.003 | +0.7% | 0.059 | +4.4% |
| 100mV/d | 0.010 | 0.0% | 0.100 | 0.0% |
| 120mV/d | 0.022 | -1.2% | 0.147 | -5.5% |

The integrity of the Eq. (2) approach to background correction in general arises from the fact that ($I_{k\_VG} - I_{k\_VC}$ ) is identically equal to the measured current difference ($I_{k\_VG\_VC} - I_{all\_VC}$) and the error in, for example, $I_{off}$ can in no case be worse than the measurement error in ($I_{k\_VG\_VC} - I_{all\_VC}$) combined with an error associated with the variation in $\eta_k$ , This latter error can have a fractional value no larger than the actual value of $\eta_k$ . Furthermore with $\eta_{all}$ = 0.02 and a 30 DUT array, $I_{off}$ has a value in the same range as $I_{all\_VC}$ so the instrument related measurement accuracy in determining ($I_{k\_VG\_VC} - I_{all\_VC}$) for similar MOSFETs is of the same order of that in directly measuring $I_{off}$ values of isolated MOSFETs. For heterogeneous DUTs, if $I_{all\_VC}$ becomes large compared to $I_{off}$ , instrument related error will ultimately increase. Although as long as the range of $I_{off}$ for the set of DUTs due to, for example, differences in channel lengths, channel widths and $V_t$ values is of order 30 or less that error will still generally not be large.

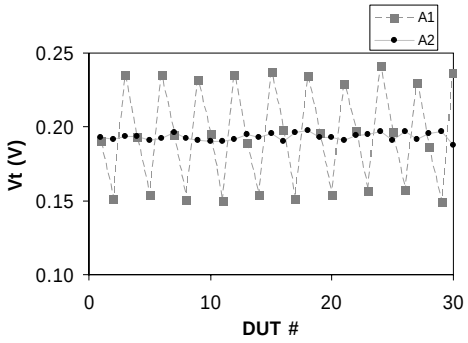



Figure 5: Mean $V_t$ vs. DUT# of 64 chips on a wafer for two arrays, A1 with three different $V_t$'s by column and A2 with a nominally same $V_t$ for all DUTs.

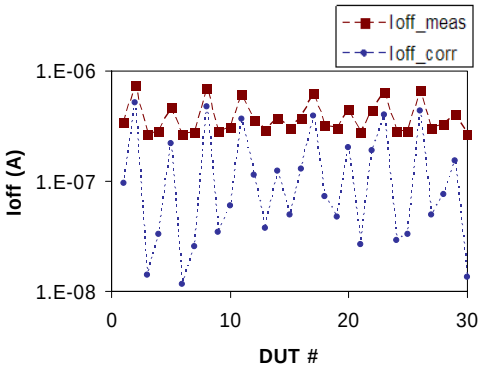



Figure 6: Measured $I_{off\_meas}$ and corrected $I_{off\_corr}$ vs. DUT# for the array A1 on a single chip.

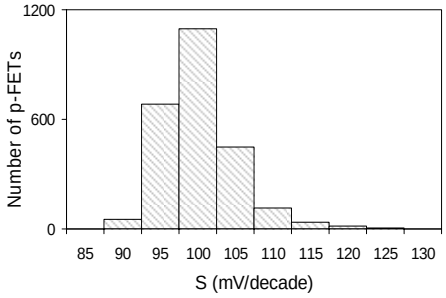



Figure 7: Distribution of sub-threshold slope, S, for a population of p-FETs.

For ultra-low power technologies with high threshold voltages, GIDL can be a significant part of $I_{off}$. In such cases the methodology of Eq. (2) will not be appropriate for determining sub-threshold currents in the vicinity of $I_{off}$.

Also the gate dielectric tunneling in the MOSFET DUTs can become a concern in the sub-threshold regime for sufficiently thin gate dielectrics and/or sufficiently low source-drain channel currents. With the current measured at the drain terminal, any gate dielectric tunneling current will tend to add to the drain current. The background correction given in Eq. 2 will continue to hold, although $I_{k\_VG}$ will now include a component of the selected DUT's gate tunneling current. Note that the gate current is expected to increase as |VC| is increased, as opposed to decreasing as the source-drain channel current does. Thus in determining the value of $\eta_{all}$ and establishing an optimum value for VC, gate dielectric tunneling may contribute to the behavior attributed to GIDL. On the other hand provided VC can be set so that $\eta_{all} << 0.1$, it follows that gate dielectric tunneling is not a significant contributor to $I_{k\_VG}$ for VG > 0. For the high performance 45 and 32 nm technologies in which we implemented these arrays, this is typically the case. Looking to the future the migration to high-k gate dielectrics will diminish the impact of such gate dielectric tunneling.

The MOSFET test structure described here can be reconfigured to accurately measure the gate dielectric tunneling currents. In this redesign, the source pad is isolated from the decoder ground(n-FET) or VDD (p-FET) and currents can be measured at both the source and the drain terminals or with the source and drain connected together. However, the redesign consumes additional I/O's and reduces the number of array-units in a macro.

The total $V_t$ variation for a population of 1920 nominally identical FETs on a wafer is shown in Fig. 8. Also shown is the systematic $V_t$ variation as determined from the mean $V_t$ from the 30-DUT array measured on each of the 64 chips. The distribution of the mean values excludes the intrinsic device variability and quantifies the across wafer variations for process tuning.

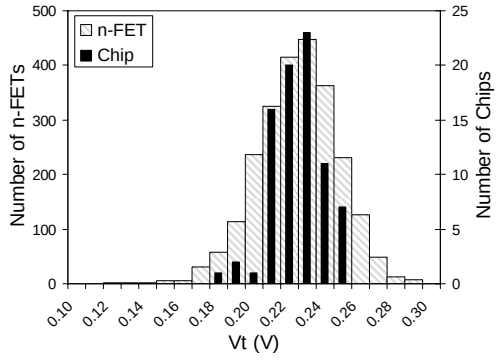



Figure 8: Histograms showing total $V_t$ variation for 1920 n-FETs on a wafer and $V_t$ variation of the array mean for an array of 30 n-FETs on each of 64 chips.

The resistor array macro is configured for measuring metal wires, polysilicon segments, and contact via resistances. The GND resistance as a function of DUT location was measured in a resistor array-unit in which all DUTs were shorts to GND. The measured GND resistance as well as simulated values from full-parasitic extraction and SPICE simulations are shown in Fig. 9 and correlate well. As previously mentioned, DUTs #1 and #30 in each array can be shorts to GND and a resistance correction applied using a linear fit. The current through the DUT is < 1 mA for negligible power supply voltage droop, and measurements with an accuracy of ± 0.2 Ohm can be made with a standard parametric tester. In Fig. 10, resistance in arbitrary units (a.u.) vs. DUT # for metal wire segments and metal-Si contacts in arrays are shown, where the measured values have been corrected for the ground resistance as indicated in Fig. 9. In Fig. 11, the distribution of corrected metal-Si contact resistances on a wafer is shown. In this case, the long tail in the distribution related to a processing error is clearly visible.

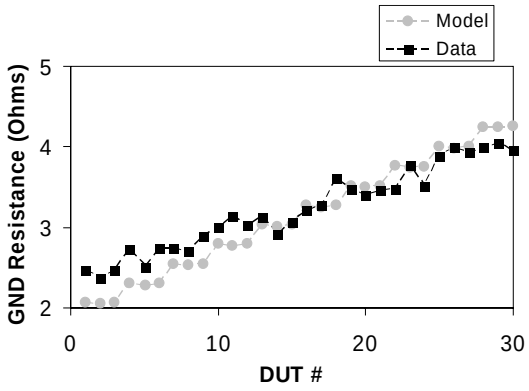



Figure 9: Simulated and measured resistance vs. DUT# in an array with all DUTs shorted to GND.

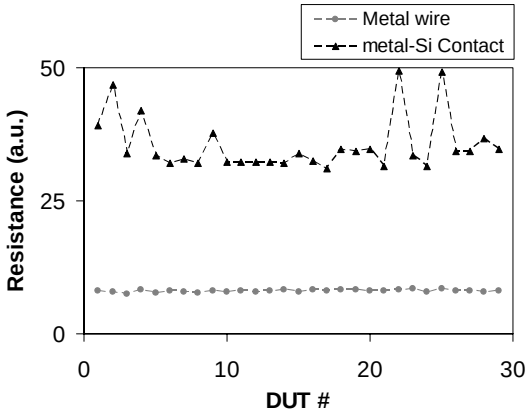



Figure 10: Resistance vs. DUT# of 29 nominally identical metal wire segments and 29 nominally identical metal-Si contacts.

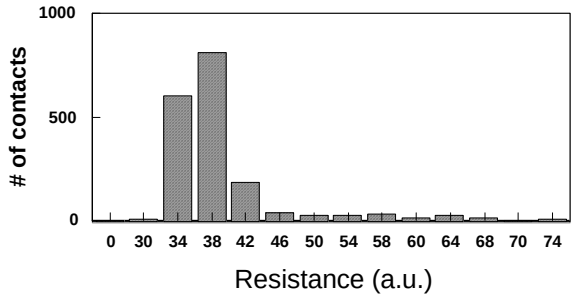



Figure 11: Distribution of metal-Si contact resistance on a wafer showing a long high resistance tail region.

## SUMMARY

Test structures for measuring arrays of MOSFETs and parasitic resistances have been implemented with a single level of metal. These compact test structures can be placed in the scribe line and are suitable for collecting modest statistics from early technology learning to product manufacturing. The full range of I-V characteristics including the sub-threshold region can be obtained as long as gate dielectric leakage and GIDL are small. Resistances of elements such as metal-Si contacts, metal wire segments and polysilicon segments can be measured with ± 0.2 Ohm precision with a standard parametric tester.

## ACKNOWLEDGEMENT

The authors gratefully acknowledge the contributions of Anne Gattiker in data analysis and the support of many individuals in the IBM Systems and Technology Group in the implementation of these test structures.

## REFERENCES

[1] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G.Slavcheva, ‘‘Simulations and Intrinsic Parameter Fluctuations in Decananometer and Nanometer-Scale MOSFETS,’’ IEEE Trans. Electron Devices Vol. 50, p. 1838, 2003.

[2] T. Mizuno et. al.,“Experimental Study of Threshold Voltage Fluctuations Using an 8k MOSFET Array”, Symp. VLSI Tech. Digest, pp.41-42 (1993).

[3] S. B. Yeo, J. Bordelon, S. Chu, M. F. Li, B. A. Tranchna, M. Harward, L. H. Chan, A. See, “A Robust and Production Worthy Addressable Array Architecture for Deep Sub-micron MOSFET’s Matching Characterization”, Proceedings of the IEEE International Conference on Microelectronic Test Structures, 2002, pp. 229-234.

[4] Yoshiyuki Shimizu, Mitsuo Nakamura, Toshimasa Matsuoka and Kenji Taniguchi, “Test Structure for Precise Statistical Characteristics Measurement of MOSFETs”, Proceedings of the IEEE International Conference on Microelectronic Test Structures, 2002, pp. 49-52.

[5] Shin-ichi Ohkawa, Masakazu Aoki, and Hiroo Masuda, “Analysis and Characterization of Device Variations in an LSI Chip Using an Integrated Device Matrix Array”, IEEE Trans. on Semiconductor Manufacturing, Vol. 17, pp. 155-165, 2004.

[6] Naoki Izumi, Hiroji Ozaki, Yoshikazu Nakagawa, Naoki Kasai, and Tsunetoshi Arikado, “Evaluation of Transistor Property Variations Within Chips on 300-mm Wafers Using a New MOSFET Array Test Structure”, IEEE Trans. on Semiconductor Manufacturing, Vol. 17, pp. 248-254, (2004).

[7] Ulrich Schaper, Jan Einfeld, and Anke Sauerbrey, “Parameter Variation on Chip Level”, Proceedings of the IEEE International Conference on Microelectronics Test Structures, 2005, pp.155-158.

[8] Kanak Agarwal, et. al., “A Test Structure for Characterizing Local Device Mismatches”, Symp. VLSI Tech., pp. 129-130, (2006).

[9] Kelvin Y.Y. Doong, et. al., "Field-Configurable Test Structure Array(FC-TSA): Enabling design for monitor, model and manufacturability", Proceedings of the IEEE International Conference on Microelectronics Test Structures, 2006, pp. 98-103.

[10] Akira Mizumura, Tetsuya Oishi, Nobuhiko Yokoyama, Mie Nonaka, Shigeyuki Tanaka, and Hiroaki Ammo, “Study of 90-nm MOSFET Subthreshold Hump Characteristics Using Newly Developed MOSFET Array Test Structure”, IEEE Trans. on Semiconductor. Manufacturing, Vol. 19, pp. 19-26, (2006).

[11] Nigel Drego, Anantha Chandrakasan, and Duane Boning. “A Test-Structure to Efficiently Study Threshold-Voltage Variation in Large MOSFET Arrays”, Proceedings of the 8th International Symposium on Quality Electronic Design (ISQED'07), 2007.

[12] Christopher Hess et. al., “Device Array Scribe Characterization Vehicle Test Chip for Ultra Fast Product Wafer Variability Monitoring”, Proceedings of the IEEE International Conference on Microelectronics Test Structures, 2007, pp. 145-149.

[13] Y. Taur and Tak Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, Cambridge, UK 1998, pp. 100-128.

# Accurate Time Constant of Random Telegraph Signal Extracted by a Sufficient Long Time Measurement in Very Large-Scale Array TEG

T. Fujisawa[1], K. Abe[1], S. Watabe[1], N. Miyamoto[2], A. Teramoto[2], S. Sugawa[1] and T. Ohmi[2]

[1] Graduate School of Engineering, Tohoku University
[2] New Industry Creation Hatchery Center, Tohoku University
Aza-Aoba 6-6-10, Aramaki, Aoba-ku, Sendai 980-8579, Japan
E-mail: fujisawa@fff.niche.tohoku.ac.jp

## ABSTRACT

To suppress Random Telegraph Signal (RTS) noise in MOSFETs, it is necessary to understand the phenomena of RTS. We can extract the accurate time constant in RTS noise by measuring a huge number of MOSFETs during a long time. Time constant is useful to obtain the energy level. In this paper, we demonstrated the statistical and accurate measurement method of the time constant of RTS by a sufficient long measuring in very large-scale array TEG.

## INTRODUCTION

As shrinking of device size in MOSFETs, the noise and variation characteristics become critical issue for realizing low power consumption, high performance and highly reliable LSI [1]. It has been reported that the RTS noise became a very crucial problem for analog devices, flash memory, and so on [2], [3]. RTS noise shows discrete and stochastic switching with two and/or more states at the drain current or threshold voltage. This dynamic phenomenon is induced by the capture and emission of carriers in individual traps near the silicon-gate insulator film interface [4], [5]. To suppress the RTS noise, it is required to determine the traps energy levels. We have reported that the statistical evaluation is essential to analyze the RTS characteristics, and we can detect a very small probability of 0.1~1% MOSFETs with large RTS noise from the 1,200,000 cells during very short time [6], [7]. To obtain the energy level of traps, we must extract accurate time constant ratio of RTS. Time constant ratio has strongly relationship with energy distribution of traps in the gate insulator. It is very difficult to extract individual time constant statistically. However, time constant ratio can be extracted relatively easier than the extraction of each time constant individually. In this paper, we demonstrate the extraction of the accurate time constant ratio of RTS by using newly developed array TEG.

## MEASUREMENT METHODS

### *A The test Structure for measuring RTS*

Fig.1 shows the circuit schematic view of newly developed test structure for measuring the electrical characteristics in a very large number of MOSFETs during very short time [8]. The test structure is simply composed of the unit cells with the measured and select switch MOSFETs (A) and simple peripheral circuits. The vertical and horizontal shift register (B, C) locate to address measured MOSFETs. The current source MOSFETs (D) locate on every column to determine the current of the measured MOSFETs. The analog memories (E) locate on every column to store source voltages of measured MOSFETs, and source follower circuit (F) to amplify output voltages. It is easy to realize high integration of this TEG because the unit cell is constructed with only two MOSFETs, which are select and measured MOSFETs. The source voltage of one measured MOSFET can be read out in 0.33μsec and 1.2 million MOSFETs, which are measured with 1 frame, can be read out in 0.72 seconds by scanning with the horizontal and vertical shift registers. This TEG can operate in a wide range, as a result, it can evaluate various MOSFETs with different gate lengths, gate widths, gate insulators, thicknesses, and so on.

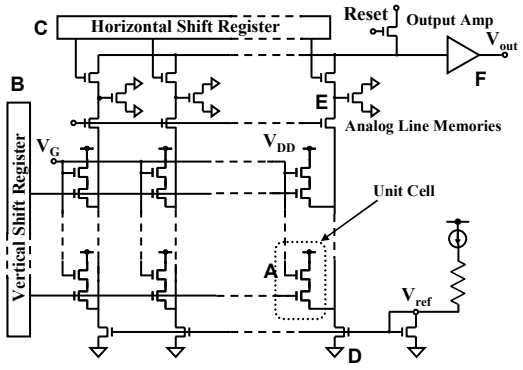



Fig. 1. The schematic view of test structure for measuring RTS.

Fig.2 (a) shows the circuit diagram of the unit cell, which is consisted measured MOSFET, switching MOSFET. A constant current ($I_{DS}$) flows when a gate voltage ($V_{ref}$) is applied to current source MOSFET. When the switch MOSFET turns on, $I_{DS}$ flow in the measured MOSFET. As a result, the voltage value between gate and source ($V_{GS}$) can be determined. On-resistance of switching MOSFET is small enough compared with the channel resistance of measured MOSFET, the source voltage of measured MOSFET equals to output voltage (Vout). Because $V_G$ is applied from outside, the $V_{GS}$ can be determinate from $V_G$-Vout. By changing the $V_{ref}$, $I_{DS}$-$V_{GS}$

characteristics can be measured. To reduce fluctuation of $I_{DS}$ induced by device dimension variation, we design the current source MOSFET with large size of W/L =15/10μm. The amplitude and appearance probability of RTS noise are reduced in the MOSFET with large size. The switch MOSFET, current source MOSFET, and other MOSFETs in peripheral circuits were designed sufficiently to avoid the additional noise during the measurement. Fig.2 (b) shows the $I_{DS}$-$V_{GS}$ characteristics in a MOSFET having RTS noise. When electron is captured or emitted, $V_{GS}$ varies during higher state and lower state, respectively. RTS amplitude is determinate by $\Delta V_{GS}$. We consider the MOSFETs with large RTS noise when large $\Delta V_{GS}$ has been observed.

Fig.3 shows the measurement system in this work. The output voltage is read out as 12-bit digital signal through the Analog Digital Converter (ADC) circuit with noise-reduction function. Field Programmable Gate Array (FPGA) on the measurement board generates driving pulses and executes digital signal processing. Low Voltage Differential Signaling (LVDS) technology has been introduced as the I/O system between the board and the data storage Personal Computer.

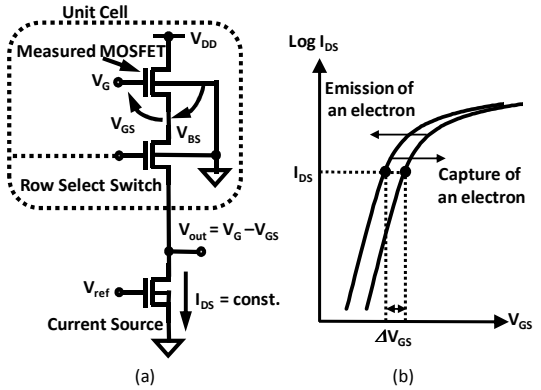



Fig. 2 (a) Circuit diagram of the unit cell, which consisted with measured MOSFET, row select switch, and current source. (b) Schematic of $I_{DS} - V_{GS}$ characteristics in a MOSFET with large RTS noise.

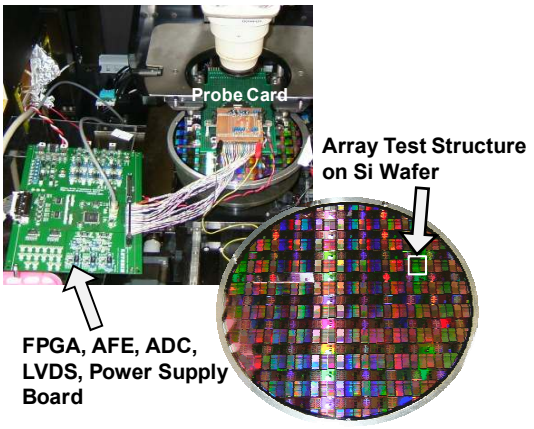



Fig. 3. Picture of measurement system.

*B Detect MOSFETs with RTS*

We have reported that MOSFETs with large RTS can be easily detected by using newly developed TEG [7]. Fig.4 is the cumulative probability distribution of the standard deviations of the continuous 300-frame outputs of each MOSFET ($\sigma_R$). $\sigma_R$ is defined as following equation;

$$\sigma_R = \sqrt{\frac{\sum_i^n \left(V_{GS_i} - \langle V_{GS} \rangle\right)^2}{n-1}}, \quad (1)$$

n and $\langle V_{GS} \rangle$ are sampling points and average $V_{GS}$, respectively. The sample has been measured with a sampling period of 0.72 sec. Temperature of measurement is 25°C. The MOSFET size is W/L = 0.28/0.22μm. Back-gate-bias is 1.0V and drain current of 0.1μA. The slope of distribution drastically changes around to the point of 99 percent. Fig.5(a), (b), and (c) show $V_{GS}$ as a function of sampling time for 3 MOSFETs. These 3 type MOSFETs are chosen from the cumulative probability distribution shown fig.4. MOSFETs shown as figs. 5(a), (b), and (c) are indicates cumulative probability of 99.998% (A), 99.97% (B) and 54.5% (C), respectively with sampling period of 0.33μsec. RTS noise is observed at the MOSFET with gradual slope. It is demonstrated that we can easily detect the MOSFETs with RTS noise from a large amount of MOSFETs by $\sigma_R$.

*C Extraction time constant*

We try to extract the amplitude and time constant ratio for analyzing the energy level of traps. Fig. 6(a) shows the $V_{GS}$ of nMOSFET with RTS as a function of sampling times of 0.33μsec with constant $I_{DS}$. $\tau_c$

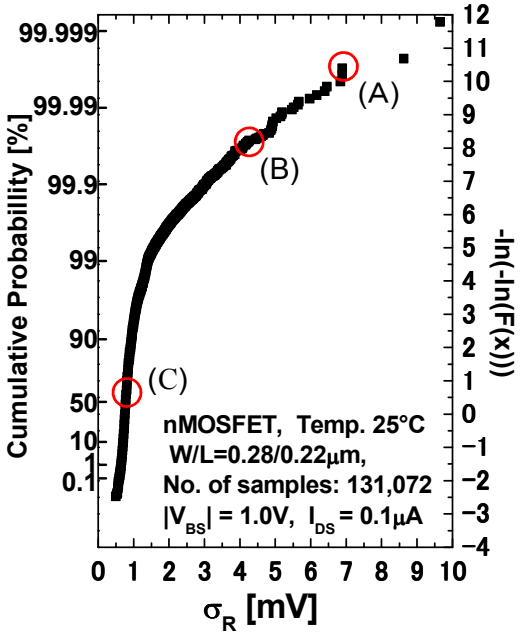



Fig. 4. Cumulative probability distribution of the standard deviation, calculating from 300 times $V_{GS}$ for one MOSFET.

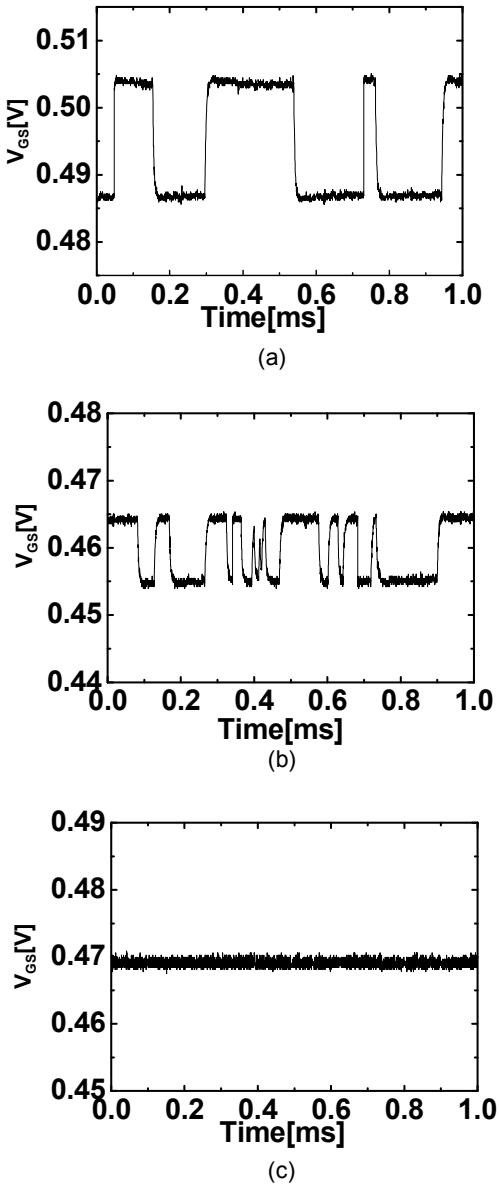



Fig. 5. (a), (b), and (c) show the $V_{GS}$ as a function of time measurement for 3 type MOSFETs which show fig. 4(A), (B) and (c), respectively.

and $\tau_e$ show time to capture and time to emission of a trap, respectively. When $V_{GS}$ exists in the higher state and the lower state, single electron is captured by oxide trap and emitted from by one, respectively. The $\tau_c$ and $\tau_e$ are determined by the sampling period when $V_{GS}$ of distribution is at the lower state and higher state respectively. Fig. 6(b) shows the frequency as a function of $V_{GS}$ from fig. 6(a). Two peaks observed at the MOSFETs with RTS have been shown at Fig.6 (b). Amplitude ($\Delta V_{GS}$) is defined as the difference between these two peaks. $V_{min}$ is defined as the $V_{GS}$ with the minimum frequency between two peaks. $V_{min}$ are used to decide lower state or higher state. CountL and countH is defined as the number of the lower state and higher state respectively. CountLH and countHL is defined as the number of transition from lower state to higher state and from higher state to lower state respectively. We measure countL, countH, countLH, and countHL from frequency of $V_{GS.}$ $<\tau_c>$ and $<\tau_e>$ show average of time to capture and time to emission respectively. $<\tau_c>$ and $<\tau_e>$ are defined as the following equations;

$$<\tau_c> = \text{countL} \times T / \text{countLH} \quad (2)$$

$$<\tau_e> = \text{countH} \times T / \text{countHL} \quad (3)$$

T shows sampling period.

Fig. 7(a) shows the $V_{GS}$ as a function of sampling number measured with the rate of 0.72 sec at constant $I_{DS}$ and (b) shows frequency of $V_{GS}$ from the result shown at (a). One MOSFET is measured with 0.33μsec sampling period one time, and 1.2 million MOSFETs are measured during 0.72sec period at once. A measurement of 1.2 million MOSFETs cannot determine the accurate time constant $<\tau_c>$ and $<\tau_e>$ because T is much longer than transition time of $V_{GS}$. But we can measure time constant ratio $<\tau_c>/<\tau_e>$ because $<\tau_c>/<\tau_e>$ is not effected with sampling period are shown as following equation which is introduced with (2) and (3):

$$\frac{\langle \tau_c \rangle}{\langle \tau_e \rangle} = \frac{\text{countL}}{\text{countH}} \times \frac{\text{countH}}{\text{countLH}}, \quad (4)$$

and $<\tau_c>/<\tau_e>$ approach to exact value even if sampling points are much sufficiently. In this work, we measure many accurate time constant ratios at once.

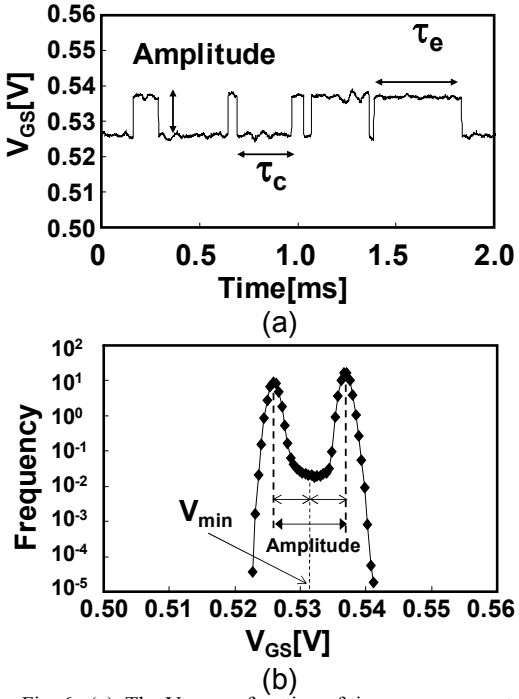



Fig. 6. (a). The $V_{GS}$ as a function of time measurement with the rate of 0.33 μsec at constant $I_{DS}$.
(b) Frequency of $V_{GS}$ from (a).

Fig. 8 shows the sequence for extracting accurate time constant of RTS. At first, we measure 1.2million MOSFETs with 6000 sampling every 0.72sec, and detect two peaks by using the distribution of $V_{GS}$. We consider that RTS is detected when the differences of two peak larger than 4.2mV, which is as large of 6 times as the noise in this measurement system, and number of second peak more than 30, which confirm whether frequency of $V_{GS}$ has two peaks or not. Seconds, we judge whether extracted time constant ratio is accurate. We measure deviation of time constant ratio with sampling number, and its probability decades less than $\pm$ 5%, we judge extracted time constant ratio is accurate.

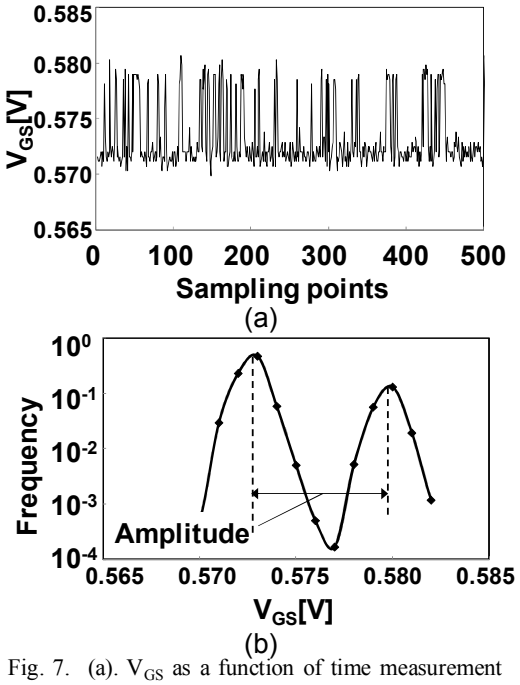



Fig. 7. (a). $V_{GS}$ as a function of time measurement with the rate of 0.72 sec at constant $I_{DS}$.
(b) Frequency of $V_{GS}$ from (a).

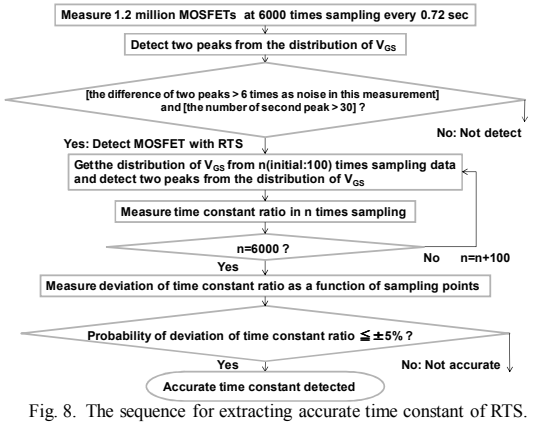



Fig. 8. The sequence for extracting accurate time constant of RTS.

## RESULTS AND DISCUSSIONS

In this work, we measure one dimension MOSFETs in TEG, which indicate 131,072 MOSFETs of W/L = 0.28/0.22μm. Fig.9 shows the deviation of time constant ratio, $(\Delta(\langle\tau_c\rangle/\langle\tau_e\rangle))/(\langle\tau_c\rangle/\langle\tau_e\rangle)$, as a function of the sampling number. The deviation of time constant ratio decrease with sampling number. Sampling period is 0.72 sec. $\Delta(\langle\tau_c\rangle/\langle\tau_e\rangle)$ represents the difference of $\langle\tau_c\rangle/\langle\tau_e\rangle$ for n times and the as (n+1) times. We judge that $\langle\tau_c\rangle/\langle\tau_e\rangle$ approaches to exact value when $(\Delta(\langle\tau_c\rangle/\langle\tau_e\rangle))/(\langle\tau_c\rangle/\langle\tau_e\rangle)$ decays within ± 0.05. In other words, we measure accurate $\langle\tau_c\rangle/\langle\tau_e\rangle$ by this method. $\langle\tau_c\rangle/\langle\tau_e\rangle$ in sample A converge at about 1,000 times sampling though one converge at 4,000 times in sample B. We consider that 4,000 ~ 6,000 sampling is needed for $\langle\tau_c\rangle/\langle\tau_e\rangle$ to obtain the exact value.

Figs.10 and Figs.11 show the relationship between Amplitude and $\langle\tau_c\rangle/\langle\tau_e\rangle$ of RTS noise. Sampling points is the number of 6000 times with the rate of 0.72 sec. We consider $\langle\tau_c\rangle/\langle\tau_e\rangle$, which are plotted with the distribution of Figs.10 and Figs.11, as the accurate time constant ratio. As it can be seen that the distribution of $\langle\tau_c\rangle/\langle\tau_e\rangle$ value indicates symmetric at center value of 1. It indicates that we can easily discover the sample which the value of $\langle\tau_c\rangle/\langle\tau_e\rangle$ are around 1 in this work. It is consider that $\langle\tau_c\rangle/\langle\tau_e\rangle$ become far from 1 when amplitude is high.

Figs.10 show the amplitude dependency of $\langle\tau_c\rangle/\langle\tau_e\rangle$ on different measurement temperatures. The MOSFETs are measured at drain current of 1.0μA, back-gate-bias of 1.0V, and the sampling points is 6000 times with the rate of 0.72 sec. The results measured at -25°C is shown at (a), -10°C is shown at (c), 25°C is shown at (c). Red points indicate that RTS is observed and $\langle\tau_c\rangle/\langle\tau_e\rangle$ converge enough even if temperature change. (d) shows the results measured at common points for (a), (b), and (c). Square, circle, and triangle show -25°C, -10°C, and 25°C, respectively. 271 points are existed in the plot for each temperature. We can measure accurate $\langle\tau_c\rangle/\langle\tau_e\rangle$ even if temperature changes from -25 °C to 25 °C in these 271 points.

Figs. 11 show the amplitude dependency of $\langle\tau_c\rangle/\langle\tau_e\rangle$

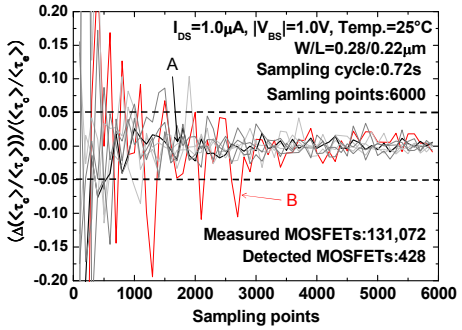
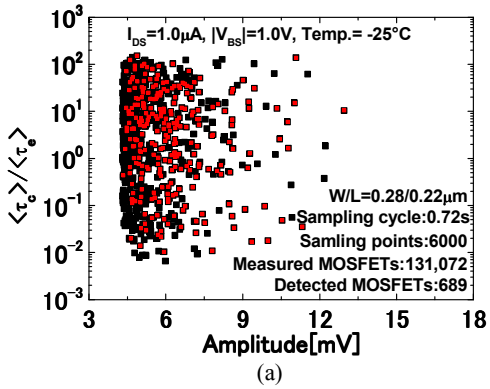
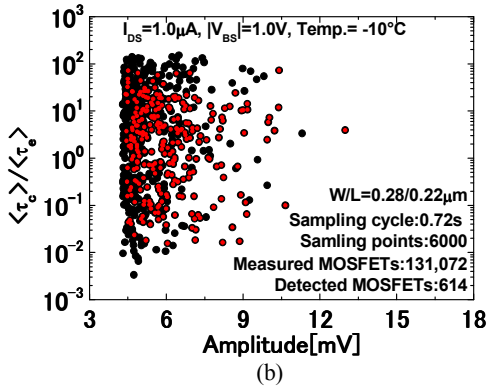


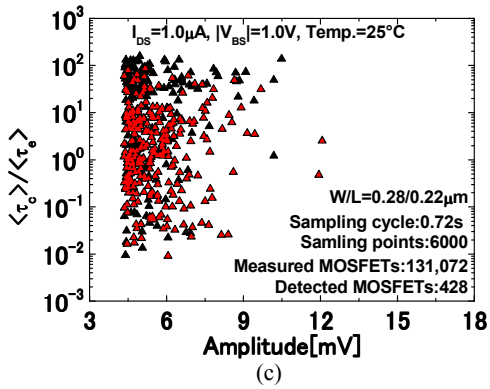

Fig. 9. Deviation of $(\Delta(\langle\tau_c\rangle/\langle\tau_e\rangle/(\langle\tau_c\rangle/\langle\tau_e\rangle))$ as a function of the sampling points. $(\Delta(\langle\tau_c\rangle/\langle\tau_e\rangle)/(\langle\tau_c\rangle/\langle\tau_e\rangle))$decrease with increase of sampling points. Sampling period is 0.72 sec.



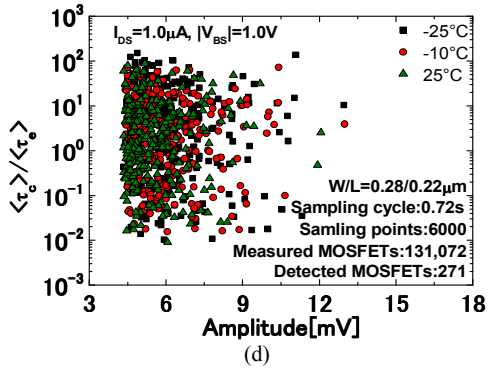

(a)




(b)

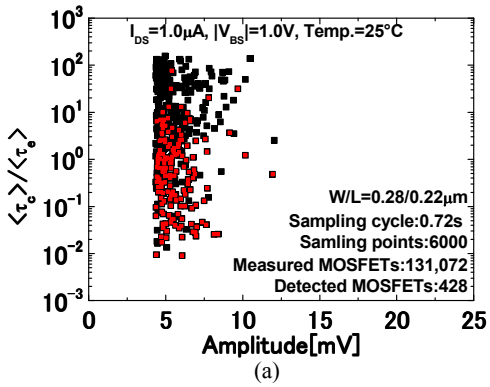



(c)

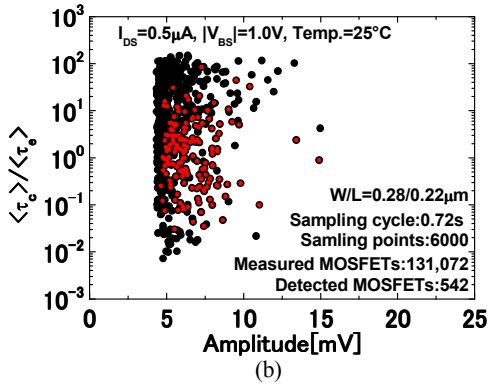



(d)

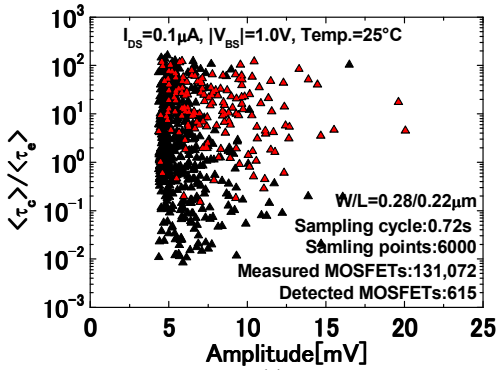
Fig. 10. Amplitude dependency of $\langle\tau_c\rangle/\langle\tau_e\rangle$ for changing of some temperature. (a) shows -25ºC, (b) -10ºC, (c) 25ºC and (d) shows common point for (a), (b), and (c).



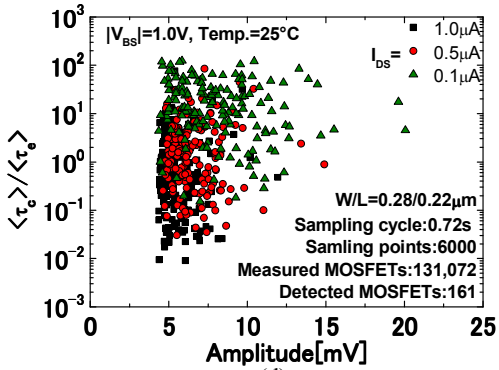

(a)




(b)




(c)

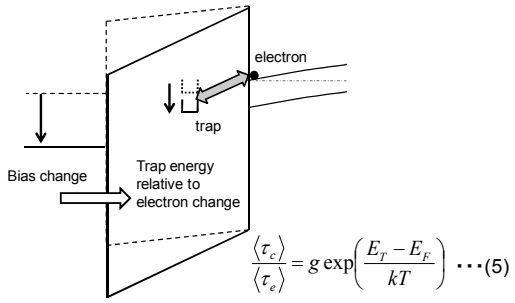



(d)

Fig. 11. Amplitude dependency of $\langle\tau_c\rangle/\langle\tau_e\rangle$ for changing of some drain current. (a) shows 1.0μA, (b) 0.5μA, (c) 0.1μA and (d) shows common point for (a), (b), and (c).

on different drain current. Temperature is 25°C, back-gate-bias is 1.0V, and sampling points are 6000 times in the rate of 0.72 sec. (a) shows 1.0μA, (b) shows 0.5μA, (c) shows 0.1μA measurements results. Red points indicate that RTS is observed and $\langle\tau_c\rangle/\langle\tau_e\rangle$ converge enough even if drain current change. (d) shows common points for (a), (b), and (c). Square, circle, and triangle show 1.0μA, 0.5μA, and 0.1μA, respectively. 161 points are existed in the plot for each drain current. We can measure accurate $\langle\tau_c\rangle/\langle\tau_e\rangle$ even if drain current change in these 161 points. (d) shows that the $\langle\tau_c\rangle/\langle\tau_e\rangle$ increase as $I_{DS}$ decrease. This result indicates that empty probability become high when drain current become small.

Fig.12 shows schematic of band diagram with RTS trap. When gate bias varies, relative energy of traps to electron changes, $\langle\tau_c\rangle/\langle\tau_e\rangle$ is represented as (5). $E_T$, $E_F$, k, T, and g show trap energy, Fermi energy, Boltzmann constant, temperature, and degeneracy factor, g equals to 1[5].




$$\frac{\langle\tau_c\rangle}{\langle\tau_e\rangle} = g\exp\left(\frac{E_T - E_F}{kT}\right) \quad \cdots(5)$$

Fig. 12. Schematic of band diagram with RTS trap. When the gate bias varies, relative energy of traps to electron changes. $\langle\tau_c\rangle/\langle\tau_e\rangle$ is represented as (5), $E_T$: trap energy, $E_F$: Fermi energy, k: Boltzmann constant, T: temperature, g: degeneracy factor equals to 1.

## CONCLUSION

We have proposed and demonstrated the novel test structure for measuring RTS parameters. The time constant of RTS is very important for analysis of carrier traps in gate insulator films which cause RTS. In this work, we demonstrated that 4,000 ~ 6,000 sampling is needed for $\langle\tau_c\rangle/\langle\tau_e\rangle$ to extract the exact value in amount of MOSFETs by the very large-scale array TEG. We demonstrated that Amplitude and $\langle\tau_c\rangle/\langle\tau_e\rangle$ are able to measured at different temperatures and drain currents in RTS. Consequently, the extraction of accurate time constant of RTS is useful to obtain the energy level and/or distance between traps and $Si/SiO_2$ interface.


## ACKNOWLEDGEMENT

We would like to thank Mr. Y. Kamata and Mr. K. Shibusawa for useful discussions and manufacture of the test structure.


## REFERENCES


[1] Tadahiro Ohmi, Masaki Hirayama, and Akinobu Teramoto, "New era of silicon technologies due to radical reaction based semiconductor manufacturing", Journal of Physics D: Applied Physics, 39, R1-R17, January 2006.

[2] X. Wang, P. R. Adri Mierop, A. J. P. Theuwissen, "Random Telegraph Signal in CMOS Image Sensor Pixels," in IEDM Tech. Dig., 2006, pp.115-118.

[3] H. Kurata, K. Otsuga, A. Kotabe, S. Kajiyama, T. Osabe, Y. Sasago, S. Narumi, K. Tokami, S. Kamohara, and O. Tsuchiya, "Random Telegraph Signal in Flash Memory: Its Impact on Scaling of Multilevel

[4] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R.W. Epworth, and D. M. Tennant, "Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1/f?) Noise," Phys. Rev. Lett., vol. 52, 1984, pp. 228-231.

[5] M. J. Kirton and M. J. Uren, "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low frequency (1/f) noise," Advances in Physics, vol. 38, 1989, pp. 367-468.

[6] 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. Phy., vol. 46, No. 4B, pp. 2054-2057.

[7] K. Abe, S. Sugawa, S. Watabe, N. Miyamoto, A. Teramoto, Y. Kamata, K. Shibusawa, M. Toita, and T. Ohmi, "Random Telegraph Signal Statistical Analysis using a Very Large-scale Array TEG with 1M MOSFETs," in VLSI Symp. Tech. Dig., 2007, pp. 210-211.

[8] S. Watabe, S. Sugawa, K. Abe, T. Fujisawa, N. Miyamoto, A. Teramoto, and T. Ohmi, "A Test Structure for Statistical Evaluation of Characteristics Variability in Very Large Number of MOSFETs" in ICMTS 2009.

# Extracting Resistances of Carbon Nanostructures in Vias

Wen Wu, Shoba Krishnan, Ke Li, Xuhui Sun, Raymond Wu, Toshishige Yamada, and Cary Y. Yang

Center for Nanostructures, Santa Clara University
Santa Clara, CA 95053, United States

## ABSTRACT

This paper describes a current-sensing technique to extract the resistances of carbon nanostructures in via interconnects. Test structures designed and fabricated for via applications contain carbon nanofiber (CNF)-metal composites embedded in silicon dioxide ($SiO_2$). Electrical characterization of single CNFs is performed using an atomic force microscope (AFM). This technique yields a metal-CNF contact resistance of 6.4 kΩ and a lowest CNF resistivity of 1.89e-4 Ω-cm.

## INTRODUCTION

Carbon nanotubes (CNTs) and CNFs exhibit high current-carrying capacity ($10^9$ A/cm$^2$) [1-3] and robust thermal and mechanical properties [4,5]. Because of their superior properties, carbon nanostructures have become attractive alternative materials for next-generation sub-30-nm on-chip interconnects. Electrical characterization of DC and AC behavior is essential for any new material to be incorporated in chip technologies. Unlike macro and micro-scale components, nanostructures require special techniques for characterization because it is difficult to directly probe them using conventional measurement techniques. Moreover, additional fabrication steps and extensive sample preparation may be needed before the nanostructures are ready for measurement. Imaging with microscopy such as scanning electron microscope (SEM) and scanning transmission electron microscopy (STEM) only provide structural and compositional information of the nanoscale components [6].

For I-V characterization of nanostructures, it is necessary to apply a low current or voltage to the device under test (DUT) in order to lower the power and minimize Joule heating [7]. The low-current (or low-voltage) specification requires a probing system for rapid, high-sensitivity, and low-noise measurements. To obtain resistance of nanostructures, typical two-point on-wafer measurements cannot differentiate between intrinsic DUT resistances and contact resistances between the DUT and pads/electrodes. Four-point test structures minimize the voltage drop across contacts, and have been widely adopted in horizontal one-dimension nanostructures [8,9]. However, it is challenging to fabricate equivalent four-point structures for vertical nanostructures such as vias due to complex 3-D integration of electrodes with the via. Therefore, finding an alternative way to extract contact resistance in a via is critically needed.

Electrical characteristics of CNT vias have been reported [10,11], but the magnitudes of the contact resistances were unknown. In this work, we propose a resistance extraction methodology for carbon nanostructures that separates the interfacial resistances from the CNF bulk resistance. Using an AFM on a CNF array, current distribution and surface topographies of the sample are obtained simultaneously. This allows the users to probe individual CNFs precisely and measure their electrical characteristics.

## CNF ARRAY TEST STRUCTURE

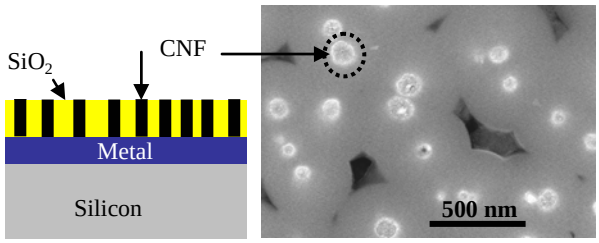



Fig. 1. (a) Cross-section of CNF arrays encapsulated in $SiO_2$, and (b) the SEM image of the sample surface, where one CNF protruding out of the oxide is highlighted by a dashed circle.

Fig. 1 (a) illustrates the cross-section of a CNF array used for electrical characterization. First we deposit a 20 nm nickel (Ni) catalyst layer on a silicon (100) wafer which is coated with a 30 nm titanium (Ti) electrode layer. Both Ni and Ti layers are deposited by ion beam sputtering. Ni catalyst film was annealed in ammonia ($NH_3$) to form Ni particles, prior to CNF growth by plasma-enhanced chemical vapor deposition (PECVD). We used acetylene ($C_2H_2$) as the carbon feedstock and $NH_3$ as the carrier gas. The inherent vertical alignment of CNFs makes integration into planar silicon processing technologies feasible. The CNF arrays are embedded in tetraethylorthosilicate (TEOS) oxide to achieve structural rigidity and electrical isolation. We polish the sample to expose the CNFs by about 30-50 nm on the surface for AFM probing. The SEM image of the sample surface is shown in Fig. 1 (b). To form an electrical path from AFM substrate chuck to the CNF base electrode (Ti), patterned metal pads (Mo/Ti) are deposited on part of the sample top surface. Then we connect the metal pads to the AFM

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

chuck with Cu tape. The parasitic resistances induced by Mo/Ti pads and the paralleled CNFs beneath them are around 50-100 Ω which is negligible compared with large resistances (kΩ) of individual CNFs.

## AFM Current-Sensing Technique

The measurement setup for the AFM is shown in Fig. 2. Platinum (Pt) coating on the surface of the AFM probe tip (tip radius < 60 nm) forms a nanoscale electrical measurement probe. With the probe tip virtually grounded, a selectable DC bias voltage is applied to the sample, and an amplifier in the controller senses the current flow through the sample. When the probe scans the sample surface in contact mode, the surface topography and current distribution are obtained simultaneously. Fig. 3 shows the 3μm x 3μm AFM images of one CNF sample. By applying a constant voltage between the probe tip and the sample base electrode, the current through several single CNFs protruding out of the oxide is obtained.

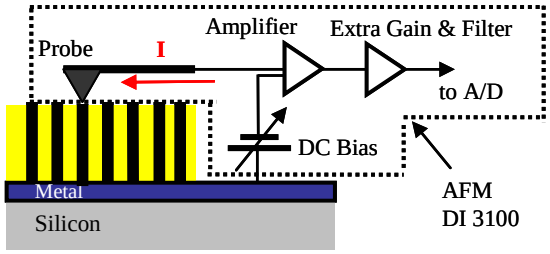



Fig. 2. AFM current-sensing measurement setup.

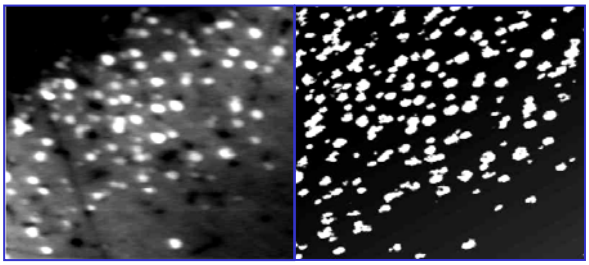


Fig. 3. (a) Surface topographic image and (b) the corresponding current distribution of one CNF sample. The white dots in (b) denote the locations of CNFs. The scan size is 3μm x 3μm.

## DC Measurements

Based on the AFM images obtained (Fig. 3), we reduce the scan size, and systematically locate single CNFs with the probe tip on the top surface of the CNF. We conduct *I-V* measurements by sweeping the DC bias voltages between -3 to +3 mV. Fig. 4 shows a linear I-V behavior obtained for one particular CNF of diameter $D_{CNF}$ = 120 nm.

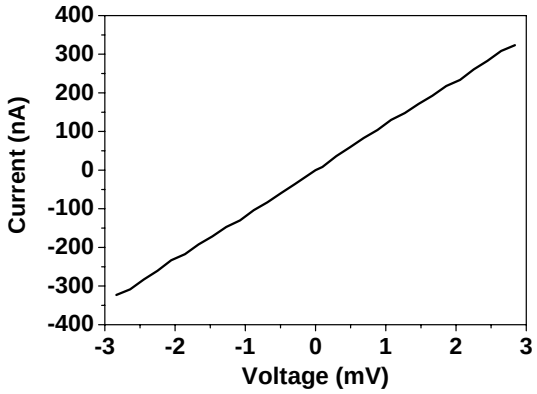



Fig. 4. *I-V* characteristics of single CNF.

## Parameter Extraction

The resistance ($R_{Total}$) inferred from the measured *I-V* curves consists of the CNF bulk resistance ($R_{CNF}$) in series with the total contact resistance ($R_C$). The following equation holds for $R_{Total}$.

$$R_{Total} = R_C + R_{CNF}. \quad (1)$$

For a CNF, when the length ($L_{CNF}$) is larger than the mean free path ($\lambda_{CNF}$), $R_{CNF}$ can be calculated from geometrical parameters as follows.

$$R_{CNF} = \frac{4\rho L_{CNF}}{\pi D_{CNF}^2} \quad (L_{CNF} > \lambda_{CNF}), \quad (2)$$

where $\rho$ is the resistivity of the CNF, and $D_{CNF}$ is the diameter of a CNF. For each sample, $L_{CNF}$ is very nearly constant.

$R_C$ represents the sum of CNF-metal and probe tip-CNF contact resistances. The contact resistance between the Pt probe tip and the CNF is dependent on the pressure applied on the tip, which can be minimized empirically. In macro-scale, contact resistance is expected to be inversely proportional to the contact area. For carbon nanostructures, this contact area or diameter dependence is more complicated [12-14]. Previous study [15] showed that tunneling is the dominant transport mechanism between a CNT tip and an electrode, which is similar to the present CNF-metal interface. If the tunneling barrier thickness $w$ has a certain correlation with $D_{CNF}$ such that $w$ is thicker for larger $D_{CNF}$, then $R_C$ can be independent of $D_{CNF}$. This is because tunneling resistance is proportional to $4/(\pi D_{CNF}^2)\exp(2\alpha w)$, where $\alpha$ is the decay constant of an electron wave function in the barrier [15]. If $\alpha w \sim \ln(D_{CNF})$, then the tunneling resistance can be almost constant. When the interface asperity is significant, as expected for the as-grown contact between

CNF and electrode, the effective $w$ for tunneling tends to be larger for larger $D_{CNF}$, making the constant $R_C$ assumption a reasonable starting point for our data analysis. Thus, $R_{Total}$ in (1) becomes a linear function of $1/(D_{CNF})^2$. Plotting $R_{Total}$ versus $1/(D_{CNF})^2$ and extrapolating to zero yields $R_C$.

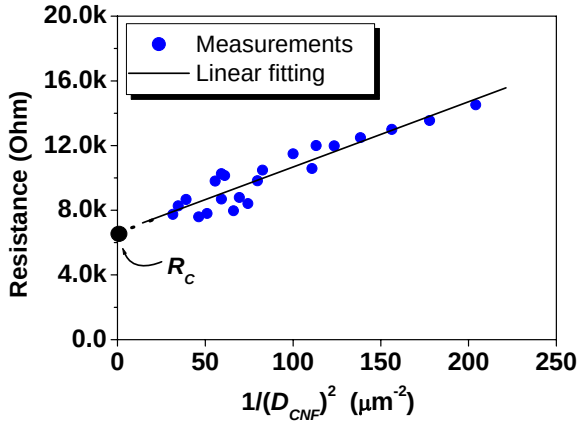



Fig. 5. Fitted curve of measured resistances versus $1/(D_{CNF})^2$. $R_C$ is the y-intercept.

*I-V* measurements were carried out on twenty-two 4.5 μm-long CNFs of varying diameters. These measurements are plotted as a function of $1/(D_{CNF})^2$ in Fig. 5. The extracted $R_C$ is 6.4 kΩ, with an error bar of about 400Ω. It is found that the lowest resistivity among measured CNFs is 1.89e-4 Ω-cm. Unlike the low-resistance behavior exhibited by metal contacts, $R_C$ values obtained here are several times larger than the bulk resistances of carbon nanostructures with similar dimensions. The contact resistance therefore dominate the electrical characteristics of carbon nanostructures in via interconnects.

For extraction of nanostructure bulk and contact resistances from measured total resistance, the method here can be extended to extract resistances from measured resistances versus varying lengths. In this case, two CNFs with similar diameters will have similar contact resistances if process variations are ignored. The limitations of this extraction method are that we need to prepare several samples with different lengths. Besides, process variations must be minimized because we assume that the difference in measured resistances of the two CNFs only results from the length difference. Study of this extraction technique using measured resistance versus length is ongoing.

## CONCLUSIONS

In this work, electrical characterization of carbon nanostructures in via interconnects has been conducted using an AFM. The current-sensing technique enables us to probe individual CNFs and obtain *I-V* curves for single nanofibers. The CNF bulk resistivity and contact resistance are then extracted from the total resistance measurements. This method is applicable to other one-dimensional nanostructures. Our results show that the contact resistances at the metal-carbon interface dominate the electrical characteristics of carbon nanostructures in via interconnects.

## ACKNOWLEDGEMENT

This work is supported by the United States Army Space and Missile Defense Command (SMDC) and carries Distribution Statement A, approved for public release, distribution unlimited.

## REFERENCES

[1] Z. Yao, C. L. Kane, and C. Dekker, "High-field electrical transport in. single-wall carbon nanotubes", *Phys. Rev. Lett.*, vol. 84, March 2000, pp. 2941-2944.

[2] B. Q. Wei, R. Vajtai, and P. M. Ajayan, "Reliability and current carrying capacity of carbon nanotubes", *Appl. Phys. Lett.* vol. 79, August 2001, pp. 1172-1174.

[3] A. Javey, P. Qi, Q. Wang, and H. Dai, "Ten- to 50-nm-long quasi-ballistic carbon nanotube devices obtained without complex lithography", *Proc. Natl. Acad. Sci. USA*, vol. 101, September 2004, pp. 13408-13410.

[4] Q. Ngo, T. Yamada, M. Suzuki, Y. Ominami, A. M. Cassell, J. Li, M. Meyyappan, and C. Y. Yang, "Structural and electrical characterization of carbon nanofibers for interconnect via applications", *IEEE Trans. Nano.*, vol. 6, November 2007, pp. 688-695.

[5] M. Suzuki, Y. Ominami, Q. Ngo, C.Y. Yang, A.M. Cassell, and J. Li, "Current-carrying capacity of carbon nanofiber interconnects", *J. Appl. Phys.*, vol. 101, June 2007, pp. 114307-114311.

[6] *Nanotechnology Measurement Handbook: A Guide to Electrical Measurements for Nanoscience Applications*, 1st Edition, Keithley Instruments, Inc.

[7] M. A. Kuroda and Jean-Pierre Leburton, "Joule heating induced negative differential resistance in freestanding metallic carbon nanotubes", *Appl. Phys. Lett.*, vol. 89, September 2006, 103102.

[8] A. Bachtold, M. Henny, C. Terrier, C. Strunk, and C. Schonenberger, J. P. Salvetat, J. M. Bonard, and L. Forro, "Contacting carbon nanotubes selectively with low-ohmic contacts for four-probe electric measurements", *Appl. Phys. Lett.*, vol. 73, July 1998, pp. 274-276.

[9] A. S. Walton, C. S. Allen, K. Critchley, M. L.

Gorzny, J. E. McKendry, R. M. D. Brydson, B. J. Hickey, and S. D. Evans, "Four-probe electrical transport measurements on individual metallic nanowires", *Nanotechnology*, vol. 18, February 2007, 065204.

[10] A. Kawabata, S. Sato, T. Nozue, T. Hyakushima, M. Norimatsu, M. Mishima, T. Murakami, D. Kondo, K. Asano, M. Ohfuti, H. Kawarada, T. Sakai, M. Nihei, and Y. Awano, "Robustness of CNT via interconnect fabricated by low temperature process over a high-density current", in *proceedings of IEEE 2008 International Interconnect Technology Conference*, pp. 237-239.

[11] J. C. Coiffic, M. Fayolle, H. Le Poche, S. Maitrejean, and S. Olivier, "Realization of via interconnects based on carbon nanotubes", in *Proceedings of IEEE 2008 Int. Interconnect Tech. Conf.*, pp. 153-155.

[12] Q. Ngo, S. Krishnan, A. M. Cassell, Y. Ominami, J. Li, M. Meyyappan, and C. Y. Yang, "Electrical characterization of carbon nanofibers for on-chip interconnect applications", in *proceedings of IEEE-Nano 2005*, pp. 431-434.

[13] Y. Massoud and A. Nieuwoudt, "Accurate resistance modeling for carbon nanotube bundles in VLSI interconnect", in *proceedings of IEEE-Nano 2006*, pp.288-291.

[14] A. Nieuwoudt and Y. Massoud, "Evaluating the impact of resistance in carbon nanotube bundles for VLSI interconnect using diameter-dependent modeling techniques", *IEEE Trans. Electron Devices*, vol. 53, October 2006, pp. 2460-2466.

[15] T. Yamada, "Modeling of electronic transport in scanning tunneling microscope tip-carbon nanotube systems", *Appl. Phys. Lett.*, vol. 78, March 2001, pp. 1739-1741.

# Demonstration of a Sub-micron Damascene Cu/Low-*k* Mechanical Sensor to Monitor Stress in BEOL Metallization

C. J. Wilson[1,2], K. Croes[2], Zs. Tőkei[2], G. P. Beyer[2], A. B. Horsfall[1], and A. G. O'Neill[1]

[1] School of Electrical, Electronic, and Computer Engineering, Newcastle University, Newcastle upon Tyne, U.K.
[2] IMEC, Kapeldreef 75, B-3001 Leuven, Belgium

## Abstract

This work reports the results of a mechanical sensor to monitor stress in 100 nm critical dimension Cu interconnects. Existing methodology developed for larger scale Al sensors is discussed and evaluated for Cu/$SiO_2$ and Cu/Low-*k* integration schemes. New sensor release methods are then developed and the Cu sensor is demonstrated in single and dual damascene technology. We also demonstrate the sensor is sensitive to process modifications and a viable tool for monitoring stress in the Cu back end of line stack.

## Introduction

The scaling of interconnect dimensions limits the ability to monitor stress during back end of line (BEOL) fabrication. Due to its simplistic nature, the wafer curvature measurement scheme is traditionally the most widely acknowledged method for characterization of residual stress in films. Film stress is determined by converting the measured curvatures using Stoney's formula [1]. However, this technique cannot be applied to a single metal stripe. For this, a very high intensity and focused x-ray beam, only available at specialized facilities, is needed to determine the strain by precision lattice parameter measurement [2]. Micro-mechanical systems offer an alternative technique which can be used to directly measure the stress at a specific point in metallization. These systems typically rely on releasing the test material from its constraining medium. If clamped symmetrically, the structure buckles [3]. Alternatively, if the structure is anti-symmetric a rotation can occur [4]. The mechanisms of each sensor then convert this into a quantifiable movement. A comprehensive review of mechanical sensors can be found in [5,6]. Each has their respective advantages and disadvantages, however, when scaled and applied to a CMOS process, the foremost concern of which is the simplicity of design ease of manufacture.

We have previously developed an Al inline rotating beam stress sensor [7]. This structure is ideally suited to BEOL integration, made up of simple 90 degree beams, similar to that of interconnects. Our work has shown, when fabricated in Al at the micron scale, the sensor has the ability to monitor stress in interconnects [8], the dependence of stress with process conditions [9], and the stress evolution during electrical reliability testing [10]. As the semiconductor industry moves from Al to Cu metallization to minimize RC delay, this technique must be adapted to the new integration schemes. Recently, this structure has been applied to "global" level 2 um Cu interconnects embedded in $SiO_2$ [11]. In this work, we scale the sensor an order of magnitude to 100 nm feature width, and demonstrate its extendibility to low-*k* stacks for the 32 nm technology node and beyond.

## Micro-Rotating Sensor

Fig. 1 shows the micro-rotating sensor used in this work; which consists of a pointer and two offset expansion arms. The stress is relieved by a "released" (detached) section of the sensor which is allowed to relax. This release is the critical step for the sensor operation [8]. As the stress is relaxed the small deviation in arm length exerts a torque about the centre of the sensor, causing the pointer to rotate. The mechanism of the sensor then converts this into a quantifiable displacement. The key parameters for the operation of the sensor are the arm separation, $L_4$ and the arm length, $L_1$. The positioning of the sensor arms on the pointer determines the direction

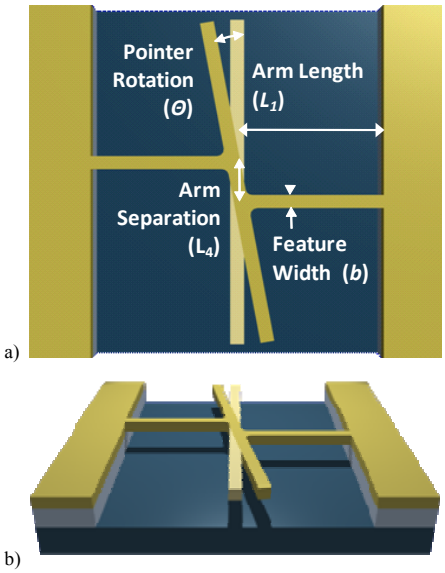



Fig. 1. Schematic representation of a released sensor depicting a tensile pre-release stress: a) top view, b) rotated side view

of rotation and sensitivity, whereas the arm length determines the maximum rotation.

Two sets of samples were used in this study to demonstrate both single and dual damascene operation. To investigate a sensor release considering the SiCN/O liner, single damascene structures were fabricated on 200 mm Si wafers, with a $SiO_2$ pre-metal dielectric (PMD), SiCN/O liner and $SiO_2$ inter-metal dielectric (IMD). Dual damascene structures were fabricated on 300 mm wafers using materials corresponding to a 32 nm technology node process; with a $SiO_2$ PMD, SiCN/O liner and a $k$=2.5 material SiOCH based IMD. This stack allows for an assessment of a sensor release in a porous low-$k$ dielectric suitable for advanced technology nodes. Identical structures were patterned in each stack with an arm-width (critical dimension) of 100 nm. In both structures the arm length was fixed at 3.5 um and the arm separation varied determine the geometrical sensitivity dependence.

## RESULTS AND DISCUSSION

Cu integration has a number of differences compared with Al, such as damascene patterning; sidewall diffusion barriers; liners; and SiOCH based ultra-low-$k$ dielectric materials, as shown in fig. 2. These modifications to the sensor fabrication will affect the sensor release. The additional layers in the damascene structure increases the number of etch steps required to "release" the sensor. Thus techniques must be optimized for etch selectivity of the various materials in both the single and dual damascene stacks, exemplified by figs. 2b and 2c respectively. Furthermore, the sensor mechanics must be tested in sub-micron features.

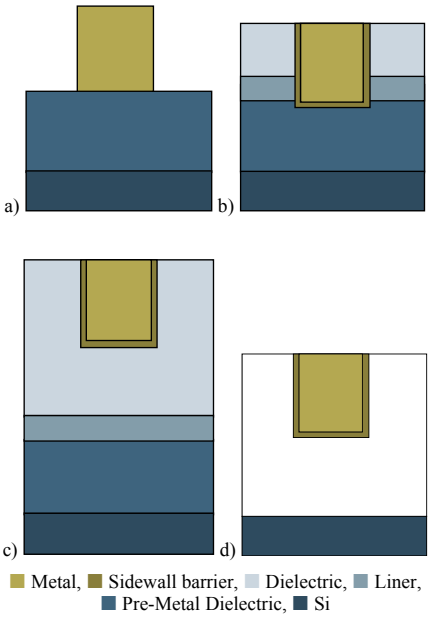



Fig. 2. Schematic representation of different integration schemes in cross-section, showing the layers/materials which must be removed for effective sensor release: a) Al metallization; b) and c) Cu single and dual damascene metallization; d) a released line

### A. *Development of Release Techniques*

A fully dry vapor hydrofluoric acid (HF) release methodology, based on the one previously developed for Al sensors [8] was trialed on various Cu stacks. A number of key disadvantages have been identified when using scaled dimensions and damascene Cu integration. HF cannot etch the SiCN/O based liners. The use of vapor HF can also leave chemical residues on the sensor thereby hindering operation. Etching of SiN with HF produces residues following $Si_3N_4 + 16HF \rightarrow 2(NH_4)2SiF_6 + SiF_4$ [12]. Such residues can be obstructive when releasing the Cu sensor, as shown in fig. 3a.

Vapor HF was also tested on the SiOCH low-$k$ materials. The materials could be removed, however, small liquid droplets were observed on the surface, as shown in fig. 3b. Dry vapor etching of phosphoric silicate glass is known to produce a liquid residue of phosphoric acid ($H_3PO_4$) [13]. It is believed the liquid residue formed during the low-$k$ etch may be carbonic acid ($H_2CO_3$). These residues cannot be removed without a chemical rinse, suggesting a fully wet chemical etch is most suitable for the SiOCH based materials, but a plasma etch must be developed for the SiCN/O based liners.

Testing of $CF_4$ based plasma chemistries has shown that the build-up of fluorine residues, as shown in fig. 3c, can

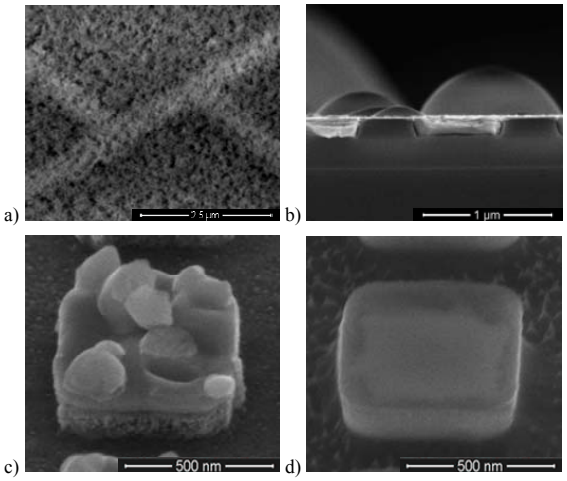



Fig. 3. Etch residues, a) a rotated top-down micrograph showing a fibrous nitride residue engulfing the sensor after dry vapor etch; b) a cross-section micrograph showing liquid residues following a dry vapor low-$k$ etch; c) a rotated top-down micrograph showing fluorine residues re-deposited on Cu following a $CF_4$ plasma chemistry and d) Cu after an $O_2$ polymer clean

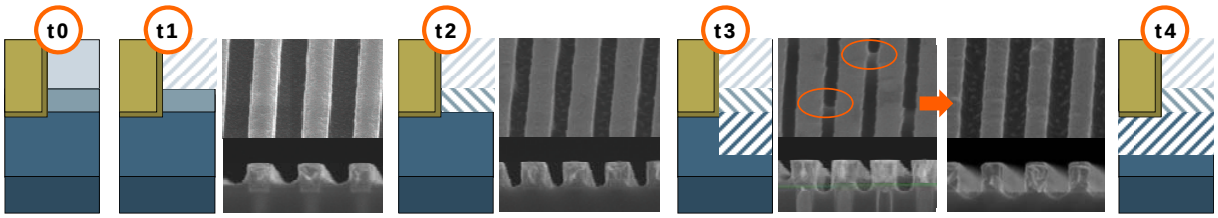



Fig. 4. Schematic illustration of the etch process using an optimized $CF_4$ chemistry: t0) initial structure, t1) dielectric etch, t2) anisotropic plasma liner each, t3) anisotropic plasma PMD etch and MS6020® clean, and t4) isotropic wet PMD etch (shaded region shows the removed material)

be catastrophic to the scaled sensors, causing severe pitting in the Cu. These were significantly reduced by introducing an $O_2$ plasma to clean the Cu during the etch, as shown in fig. 3d. However, a close-coupled polymer clean using products such as MS6020® combined with the optimized plasma may be used in more dense structures to further reduce residues.

As the wet methods trialed in this study were unable to fully etch the SiCN/O based liners present in the single damascene structure (fig. 2b), a two step approach is necessary. An optimized plasma chemistry was used to etch the dielectric and SiCN/O based liner, stopping in the PMD. This process is illustrated in fig. 4. An immersion wet etch was then used to fully release the sensor. The dual damascene structure offers a simplified release process, only requiring an isotropic wet chemical etch of one material, as shown in fig. 2c. The liquid residues observed in vapor HF trials are easily removed, causing no damage to the sensor. A similar etch process can be applied to a dual damascene Cu/$SiO_2$ structure.

### *B. Single Damascene Copper*

Fig. 5a shows a series of released sensors with varying arm separations. A change in behavior can be observed with the small changes in geometrical parameters. This is due to the balance of two forces; a bending moment about the centre of the pointer and a bending moment about the arm-pointer intersection. Unlike other work [11], the extended range of arm separations below the optimum rotation allows better fitting of the two interacting mechanisms. The angle of rotation of each sensor was extracted using a semi-automatic program, developed using the MATLAB® image processing tool box. The normalized data is plotted in fig. 6 using a 95% confidence interval. An optimum/peak rotation is clearly identifiable. Such behavior is consistent with our previous work in Al [7-10].

### *C. Dual Damascene Copper*

Fig. 5b shows a series of released dual damascene sensors with varying arm separations fabricated in a $k$=2.5 ultra-low-$k$ material. An identical change in behavior can be observed compared to the single damascene structures. The rotation data is plotted in fig. 6, showing an equivalent peak position when normalized to a trend curve. This suggests the additional release steps used for the single damascene release do not add any nonlinear effects.

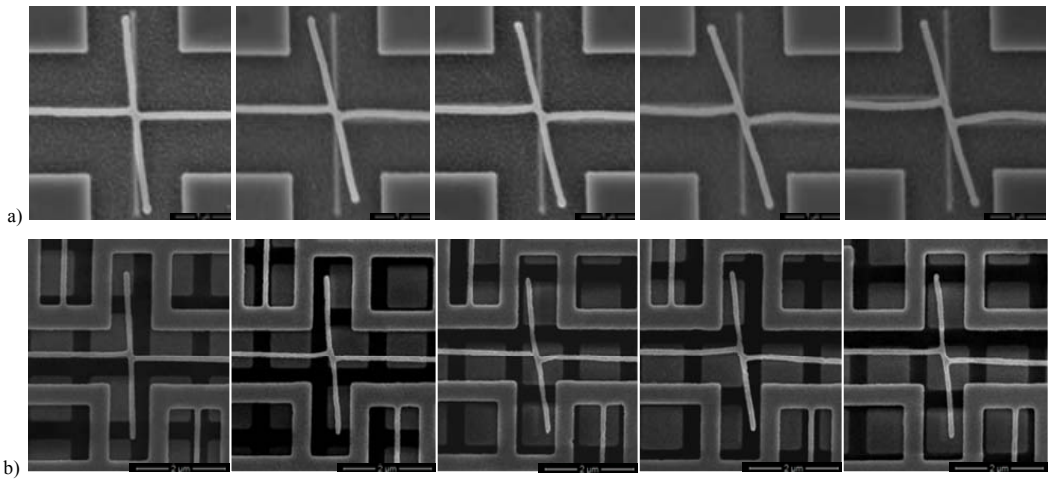



Fig. 5. Typical released Cu sensors for various arm separations patterned with 100 nm feature widths, depicting a tensile pre-release stress: a) single damascene; and b) dual damascene

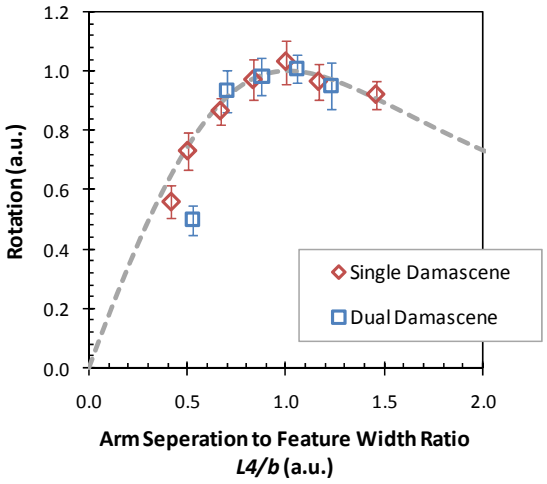



Fig. 6. Measured rotation as a function of the arm separation to feature width ratio, plotted with a 95% confidence interval

### *D. Process Monitoring*

We varied the metallization processing conditions of three wafers to alter the stress state of the Cu. Fig. 7 shows the rotations of a series of single damascene sensors. We observe the same peak/optimum interaction as the bending moment on the pointer and bending moment on the hinge are balanced. Each wafer split shows a clear change in peak rotation as a result of the different process conditions. The change in rotation is analogous to a change in the tensile stress state of the sensor arms. This demonstrates the scaled Cu sensor is sensitive to process variations and able to monitor the stress of the on chip interconnects in a similar manor to our previous work on Al [9].

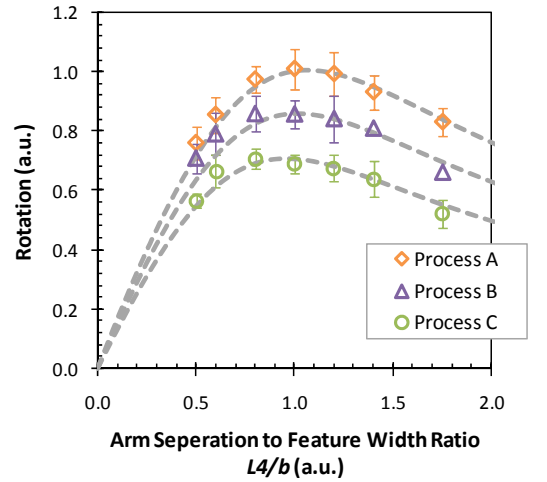



Fig. 7. Measured rotation as a function of the arm separation to feature width ratio for wafers processed with different metallization treatments, plotted with a 95% confidence interval

## SUMMARY

We have demonstrated a scaled mechanical sensor as a viable tool for monitoring stress in the Cu BEOL stack. Single and dual damascene results have been presented at feature sizes scaled to 100 nm. Conventional wet and dry etch techniques have been applied to Cu structures and new chemistries developed to minimize Cu corrosion and sensor damage. The range of sensors studied shows a characteristic geometrical behavior for the two sample sets, suggesting the additional release steps required for the single damascene release do not add any nonlinear effects. The magnitude of this curve is directly dependent on the stress in the interconnect and we have shown it can be used to monitor a given process at the "local" and "intermediate" interconnect levels. The total area required for these scaled structures allows them to be easily included between bond pads within a die, or with other monitoring structures between the scribe lines, offering a viable methodology to monitor variation in process conditions. Furthermore, the sensor has been demonstrated in a low-*k* (high-porosity) dielectric suitable for the 32 nm technology node, showing the sensor can be utilized in future integration schemes.

## ACKNOWLEDGMENTS

This work was supported by the European Commission's Information Society Technologies programme under PULLNANO project contract IST-026828; EPSRC (U.K.); and the Marie Curie APROTHIN scheme. The authors would also like to acknowledge Alexis Farcy (STMicroelectronics) and Bart Vereecke (IMEC) for preparation of test samples; Marc Van Cauwenberghe (IMEC) from the etch group for the help with the plasma work; and Kai Wang, Barry Gallacher and Steve Bull (Newcastle University) for discussions on Al/$SiO_2$ and Cu/$SiO_2$ sensors.

## REFERENCES

[1] P. A. Flinn, D. S. Gardner, and W. D. Nix, "Measurement and Interpretation of stress in aluminum-based metallization as a function of thermal history," *IEEE Trans. Electron Dev.*, vol. 34, pp. 689-699, 1987.

[2] P. R. Besser, S. Brennan, and J. C. Bravman, "An x-ray method for direct determination of the strain state and strain relaxation in micron-scale passivated metallization lines during thermal cycling," *J. Mater. Res.*, vol. 9, pp. 13-24, 1994.

[3] S. A. Smee, M. Gaitan, D. B. Novotny, Y. Joshi, and D. L. Blackburn, "IC Test Structures for Multilayer Interconnect Stress Determination," *IEEE Electron Dev. Lett.*, vol. 21, pp. 12-14, 2000.

[4] L. Lin, A. P. Pisano, and R. T. Howe, "A micro strain gauge with mechanical amplifier," *J. Micro-Electro-Mechanical Sys.*, vol. 6, pp. 313-321, 1997.

[5] B. P. Van Drieënhuizen, J. F. L. Goosen, P. J. French, and R. F. Wolffenbuttel, "Comparison of techniques for measuring both compressive and tensile stress in thin films," *Sensors and Actuators A: Physical*, vol. 37-38, pp. 756-765, 1993.

[6] N. D. Masters, M. P. de Boer, B. D. Jensen, M. S. Baker, and D. Koester, "Side-by-Side Comparison of Passive MEMS Strain Test Structures under Residual Compression," *Mechanical Properties of Structural Films*, vol. STP 1413, pp. 1-33, 2001.

[7] A. B. Horsfall, J. M. M. d. Santos, S. M. Soare, N. G. Wright, A. G. O'Neill, S. J. Bull, A. J. Walton, A. M. Gundlach, and J. T. M. Stevenson, "Direct measurement of residual stress in sub-micron interconnects," *Semicond. Sci. Technol.*, vol. 18, pp. 992-996, 2003.

[8] J. G. Terry, S. Smith, A. J. Walton, A. M. Gundlach, J. T. M. Stevenson, A. B. Horsfall, K. Wang, J. M. M. d. Santos, S. M. Soare, N. G. Wright, A. G. O'Neill, and S. J. Bull, "Test chip for the development and evaluation of sensors for measuring stress in metal interconnects," *IEEE Trans. Semicond. Manuf.*, vol. 18, pp. 225-261, 2005.

[9] A. B. Horsfall, K. Wang, J. M. M. dos-Santos, S. M. Soare, S. J. Bull, N. G. Wright, A. G. O'Neill, J. G. Terry, A. J. Walton, A. M. Gundlach, and J. T. M. Stevenson, "Dependence of process parameters on stress generation in aluminum thin films," *IEEE Trans. Dev. Mater. Reliab.*, vol. 4, pp. 482-487, 2004.

[10] C. J. Wilson, A. B. Horsfall, A. G. O'Neill, N. G. Wright, S. J. Bull, J. G. Terry, J. T. M. Stevenson, and A. J. Walton, "Direct Measurement of Electromigration-Induced Stress in Interconnect Structures," *IEEE Trans. Dev. Mater. Reliab.*, vol. 7, pp. 356-362, 2007.

[11] M. Kasbari, C. Rivero, S. Blayac, F. Cacho, O. Bostrom, and R. Fortunier, "Direct Local Strain Measurement In Damascene Interconnects," presented at Mat. Res. Soc. Symp. B, San Francisco, CA, USA, 2007.

[12] B. Du Bois, G. Vereecke, A. Witvrouw, P. De Moor, C. Van Hoof, A. De Caussemaker, and A. Verbist, "HF etching of Si-oxides and Si-nitrides for surface micromachining," presented at the Sensor Tech. Conf., Leuven, Belgium, 2001.

[13] A. Witvrouw, B. Du Bois, P. De Moor, A. Verbist, C. A. Van Hoof, H. Bender, and C. Baert, "Comparison between wet HF etching and vapor HF etching for sacrificial oxide removal," presented at SPIE Conf. on Micromachining and Microfab. Proc. Tech., VI, 2000.

# Test Structure to Extract Circuit Models of Nanostructures Operating at High Frequencies

Francisco R. Madriz, John. R. Jameson, Shoba Krishnan, Xuhui Sun, and Cary Y. Yang

Center for Nanostructures
Santa Clara University
500 El Camino Real
Santa Clara, CA 95053 USA
Phone: +1-408-554-6853, Fax: +1-408-554-5474, e-mail fmadriz@scu.edu

**Abstract**

We describe a test structure optimized for studying high-frequency electrical transport in 1-D nanoscale systems. The test structure exhibits lower transmission than previously reported structures, enabling capacitances less than 1 fF to be detected in the frequency response of the nanoscale system. The scattering parameters (S-parameters) of the test structure are describable to within ±0.5dB and ±2º from 0.1 to 50 GHz using a simple lumped-element RC circuit model whose elements are all measured experimentally.

## I. Introduction

One approach to studying electrical conduction in nanostructures is to measure the two-port scattering parameters (S-parameters) of a ground-signal-ground (GSG) test structure in which the nanostructure forms part of the signal path. This approach has been used, for example, to study single-walled carbon nanotubes [1-2], multi-walled carbon nanotubes [3-5], and bundles of carbon nanotubes [6].

A common technique for extracting the frequency response of the nanostructure itself is numerical de-embedding [7], which has been used [1-6] to study conduction in carbon nanotubes at high frequency. Another technique is to build a circuit model that accurately describes the S-parameters of the test structure in the absence of the nanostructure (*i.e.*, the "open" structure), then deduce the new elements that must be added to the circuit model to describe the S-parameters obtained *after* the nanostructure has been placed in the signal path. These added circuit elements describe electrical conduction in the nanostructure itself. These two approaches, de-embedding and circuit modeling, aim to provide information about the nanostructure under test. Circuit modeling has the advantage, however, that it provides a more in-depth physical understanding of the test structure itself. This understanding is very useful, as the design of the GSG test structure is of critical importance in obtaining highly accurate measurements.

Because transmission through individual nanoscale objects (*e.g.*, a single carbon nanotube) can be rather small, it is important to minimize the loss and transmission due to the test structure itself. In addition, it is preferable for the S-parameters of the open structure to exhibit as simple a dependence on frequency as possible, so that de-embedding or circuit modeling can be performed accurately.

This paper describes a GSG test structure designed to yield lower parasitic transmission than previously reported structures [3-6], with S-parameters that vary with frequency in a simple way. Section II describes the physical structure of the test structure; section III describes its circuit model; and, section IV explains the procedure to extract the elements of the circuit model.

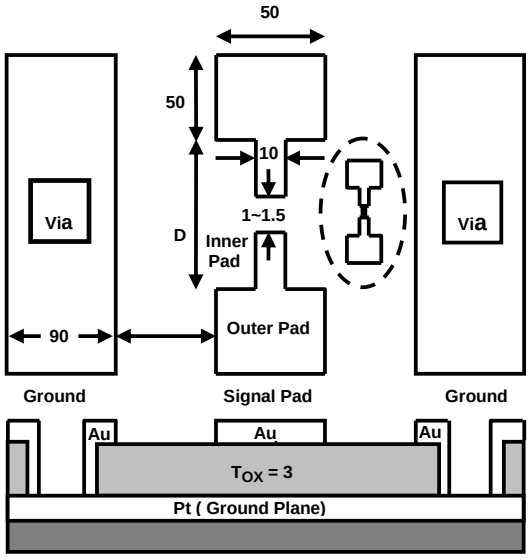



Fig. 1. High-frequency test structure layout with dimensions in µm. Drawing not to scale. (Inset) Signal pads connected by Au line of width 1 µm, used to measure $R_T$.

## II. Description of Test Structure

The physical layout of the test structure is shown in Fig. 1. Details of the fabrication process have been reported elsewhere [8]. The test structure incorporates efforts to minimize: i) transmission due to capacitive coupling between the two signal pads; ii) slow-wave propagation [9] and other losses due to coupling to the Si substrate; iii) transmission due to capacitive coupling of each signal pad to ground; and, iv) radiation between the signal pads [10]. A key feature employed to achieve i) is the use of a ground plane beneath the

$SiO_2$ layer. The ground plane also eliminates ii) [11]. To achieve iii) and iv), the thickness of the $SiO_2$ layer should be made as large as possible. However, increasing the thickness of the $SiO_2$ layer *increases* the capacitive coupling between the signal pads themselves. A trade-off thus exists between achieving i) and achieving iii-iv). Consequently, test structures with $SiO_2$ thicknesses of 1, 2, and 3 μm were fabricated and tested. Within this range of thicknesses, the largest thickness (3 μm) was found to be preferable.

### III. Circuit Model

A circuit model representing the important electrical features of the test structure is shown in Fig. 2. $R_T$ represents the contact resistance between a ground-signal-ground (GSG) RF probe tip and a signal pad. $C_{PG}$ is the capacitance between a signal pad and ground, including a contribution from the underlying ground plane, as well as small contributions from the adjacent ground pads. $R_{PG}$ represents conduction between a signal pad and ground, as well as any other loss mechanisms in the $SiO_2$ layer. $C_{PP}$ is the capacitance between the two signal pads and GSG probe tips. Finally, $R_{PP}$ represents conduction between the signal pads, and any associated losses in the $SiO_2$ layer. Other effects must exist, such as resistance and inductance in the signal pads. However, these additional effects are either made negligibly small by the geometry shown in Fig. 1, or are accounted for implicitly when the values of the circuit elements are measured experimentally.

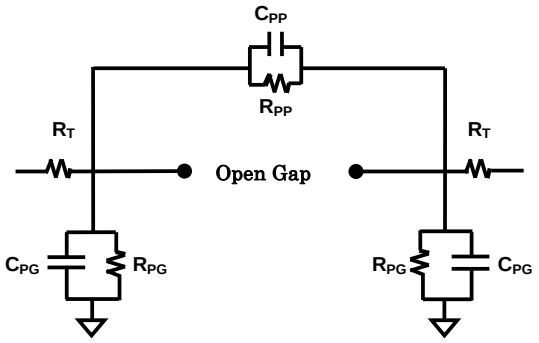



Fig. 2. Circuit model of the open test structure.

### IV. Measurement of Circuit Elements

The elements of the circuit model in Fig. 2 are measured experimentally as follows. The resistance $R_T$ is obtained using a GSG test structure in which the two signal pads are connected by a gold line 1 μm in width (see Fig. 1). A four-point measurement is performed, wherein a DC current I is forced through GSG probe tips, and the resulting voltage $V_1$ is measured. Two DC probes are simultaneously used to measure the voltage $V_2$ between the signal pads, without drawing current. The total contact resistance between the two signal pads and the two GSG probe tips is thus equal to $(V_1\text{-}V_2) / I$. Assuming the GSG probes have the same contact resistance, we obtain $R_T = (V_1\text{-}V_2) / (2I)$.

To measure $C_{PP}$ and $R_{PP}$, the two-port S-parameters of the open test structure are measured using a vector network analyzer (VNA). To calibrate the VNA, the Line, Reflect1, Reflect2, and Match (LRRM) method [12] is used, which offers higher accuracy and repeatability at frequencies above 40 GHz. The S-parameters are then converted to two-port Y-parameters using standard formulas [13]. For the parallel combination of $C_{PP}$ and $R_{PP}$, we can write

$$Y_{21} = \frac{1}{R_{PP}(\omega)} + j\omega C_{PP}(\omega) , \quad (1)$$

which readily yields $R_{PP}$ and $C_{PP}$ when equated to the measured Y-parameters. Since $R_T$ is very small compared with $R_{PP}$, it is ignored in deriving Eq. (1).

The values of $C_{PG}$ and $R_{PG}$ are determined by measuring the one-port reflection S-parameter $S_{11}$ between the signal pad and ground of a test structure that is identical to that in Fig. 1, but has had one signal pad removed. With just one signal pad, $C_{PP}$ and $R_{PP}$ do not exist, and the circuit model in Fig. 2 reduces to $R_T$ in series with the parallel combination of $C_{PG}$ and $R_{PG}$. Because $R_T$ is very small (~2 Ω), it can be neglected, and the admittance of the circuit is simply that of $C_{PG}$ in parallel with $R_{PG}$, or

$$Y_{11} = \frac{1}{R_{PG}(\omega)} + j\omega C_{PG}(\omega) \quad (2)$$

The S-parameters of the one-pad test structure are measured using a GSG probe tip and a VNA calibrated as a one-port network over its entire frequency range (0.1-50 GHz) in short, open, and load modes. In terms of $S_{11}$, $Y_{11}$ for a one-port network is defined [13] as

$$Y_{11} = \frac{1 - S_{11}}{Z_0(1 + S_{11})} , \quad (3)$$

where $Z_0 = 50\ \Omega$. Equating Eqs. (2) and (3) yields $C_{PG}$ and $R_{PG}$ in terms of the measured $S_{11}$. Fig. 3 shows the frequency dependence of $C_{PP}$ and $C_{PG}$ for the case D = 80 μm. Below ~30 GHz, the capacitance $C_{PP}$ and $C_{PG}$ are both nearly independent of frequency. Above 30 GHz, however, both decrease monotonically with frequency. This drop suggests that the test structure may contain an inductive component, perhaps from the signal pads, that has been "lumped" into $C_{PP}$ and $C_{PG}$.

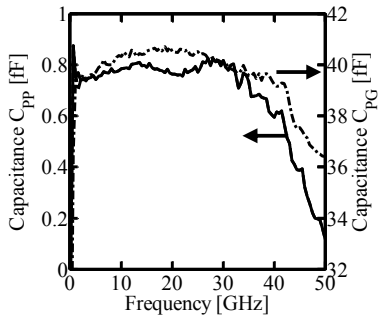



Fig. 3. Capacitance between signal pads ($C_{PP}$). Capacitance between the signal pad and ground ($C_{PG}$).

Fig. 4 compares the measured S-parameters of the open structure to the S-parameters calculated from the RC circuit model, with the values of the circuit elements obtained as described above. Excellent agreement is obtained for both the magnitude and phase of $S_{21}$ across the entire frequency range, to within ±0.5dB and ±2º from 0.1 to 50 GHz. The S-parameters exhibit fairly simple frequency dependencies, with no abrupt changes in slope. This is important because it enables changes in the S-parameters to be more easily observed when a nanostructure is placed between the signal pads. The small-scale fluctuations in the S-parameters are not random noise, but rather reflect the fluctuations in $C_{PP}$ and $C_{PG}$ shown in Fig. 3. They likely arise from many small couplings between the test structure and the measurement environment. The inset of Fig. 4 confirms that the model is successful at capturing these fluctuations.

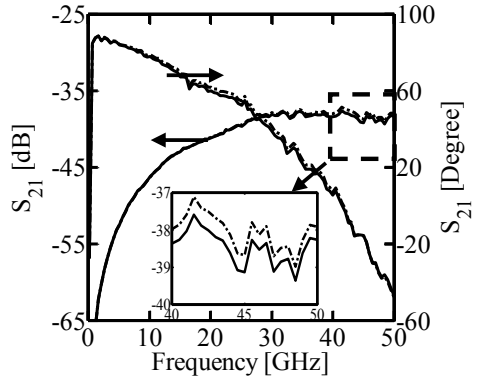



Fig. 4. Comparison of experimental (solid line) and calculated S-parameters (dotted line) for the open structure. (Inset) Enlarged view showing the agreement between model and measurement from 40 to 50 GHz.

## V. Conclusion

We have described a GSG test structure optimized for studying nanoscale systems, and have developed a circuit model that matches its two-port S-parameters very accurately. The model contains two frequency-dependent capacitances ($C_{PP}$ and $C_{PG}$), two frequency-dependent resistances ($R_{PP}$ and $R_{PG}$), and one frequency-independent resistance ($R_T$), all of which are measured experimentally. The circuit model developed matches the measured S-parameters of the open structure to within ±0.5dB and ±2º from 0.1 to 50 GHz.

## Acknowledgment

This work was supported by the United States Army Space and Missile Defense Command (SMDC) and carries Distribution Statement A, approved for public release, distribution unlimited.

## References

[1] M. Zhang, X. Huo, P. C. H. Chan, Q. Liang, Z. K. Tang, "Radio frequency transmission properties of carbon nanotubes in a field-effect transistor configuration," *IEEE Electron Lett*., vol. 29, no. 18, pp. 668-670, Aug. 2006.

[2] C. Song, Z. Liu, G. Eres, D. B. Geohegan, and P. Wang," A new method for microwave characterization of metallic single-walled carbon nanotubes," *IEEE Nano*, 2008.

[3] S. C. Jun, J. H. Cha, C. W. Baik, S. Lee, H. J. Kim, J. Hone, and J. M. Kim, "Radio-frequency transmission characteristics of a multi-walled carbon nanotube," *Nanotech.*, vol. 18, pp. 255701, May 2007.

[4] S. C. Jun, X. M. H. Huang, S. Moon, H. J. Kim, J. Hone, Y. W. Jin, and J. M. Kim, "Passive electrical properties of multi-walled carbon nanotubes up to 0.1 THz," *New J. Phys.*, vol. 9, pp. 265, Aug. 2007.

[5] G. F. Close and H. S. P. Wong, "Fabrication and characterization of carbon nanotube interconnects," *Proc. IEEE Int. Elec. Dev. Meet. (IEDM)*, pp. 203-206, 2007.

[6] J. J. Plombon, K. P. O'Brien, F. Gstrein, and V. M. Dubin, "High-frequency electrical properties of individual and bundled carbon nanotubes," *Appl. Phys. Lett*., vol. 90, 063106, Feb. 2007.

[7] R. Bauer and P. Penfield, "De-embedding and unterminating," *IEEE Trans. Microwave Theory and Techniques*, vol. 1, MTT-22, no. 3, pp. 282-288, Mar. 1974.

[8] F. R. Madriz, J. R. Jameson, S. Krishnan, K. Gleason, X. Sun, and C. Y. Yang, "Measurements and circuit model of carbon nanofibers at microwave frequencies," *IEEE Int. Interconnect Tech. Conf*., pp. 138-140, 2008.

[9] H. Hasegawa, M. Furukawa, and H. Yanai, "Properties of microstrip line on Si-$SiO_2$ system," *IEEE Trans. Microwave Theory and Techniques*, vol. 19, pp. 869-881, Nov. 1971.

[10] N. G. Alexopoulos and S.-C. Wu, "Frequency-independent equivalent circuit model for microstrip open-end and gap discontinuities," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-42, pp. 1268-1272, Jul. 1994.

[11] B. Kleveland, C. H. Diaz, D. Vook, L. Madden, T. H. Lee, and S. S. Wong, "Exploring CMOS reverse interconnect scaling in multigigahertz amplifier and oscillator design," *J. Solid-State Circ.*, vol. 36, no. 10, Oct. 2001.

[12] F. Purroy and L. Pradell," New theoretical Analysis of the LRMM Calibration Technique for Vector Network Analyzers," *IEEE Trans. Instrumentation and Measurement*, vol. 50, no. 5, Oct. 2001.

[13] D. Pozar, *Microwave Engineering*, Wiley, Massachusetts, 1998, ch. 4, pp. 210-211.

# Test Chip to Evaluate Measurement Methods for Small Capacitances

Joseph J. Kopanski, M. Yaqub Afridi, Chong Jiang[1], and Curt A. Richter

Semiconductor Electronics Division
National Institute of Standards and Technology
Gaithersburg, MD 20899-8120

**ABSTRACT**

We designed and fabricated a test chip to evaluate the performance of new approaches to the measurement of small capacitances (femto-Farads to atto-Farads range). The test chip consists of an array of metal-oxide-semiconductor (MOS) capacitors, metal-insulator-metal (MIM) capacitors, and a series of systematically varying capacitance structures directly accessible by an atomic force microscope probe. Nominal capacitances of the test devices range from 0.3 fF ($10^{-15}$ F) to 1.2 pF ($10^{-12}$ F). Measurement of the complete array of capacitances by using an automatic probe station produces a "fingerprint" of capacitance values from which, after correction for pad and other stray capacitances, the relative accuracy and sensitivity of a capacitance measurement instrument can be evaluated.

**MOTIVATION**

Due to the continual scaling of the individual devices in integrated circuits to ever smaller dimensions, the component capacitance of these nm-scale devices defy easy measurement. Emerging nanoelectronic devices fabricated from semiconductor nanowires and quantum dots, as well as double-gate fin field-effect transistor (FinFET) devices also have capacitances that are smaller than those measurable by conventional inductance-capacitance-resistance (LCR) meters. The interior device capacitances of these deep-submicron devices (such as the gate-source, source-drain, or gate-channel capacitances) determine the operational characteristics of the device, and an accurate knowledge of their values is required for accurate device modeling and predictive computer-aided design. These internal device capacitances are often difficult to directly contact and measure with probe stations and external instrumentation. Thus, an effective method to experimentally extract these capacitances will have an immediate technological impact. As an example of "small capacitance" in this context, consider calculated capacitance-voltage (C-V) curves of nanowire MOS devices with a maximum capacitance of 4 aF ($10^{-18}$ F) and a minimum capacitance of 1 aF [1]. Measurement of such a C-V curve would require better than aF accuracy.

Measurement of the small capacitances of nanoelectronic devices is inherently difficult. A single device fabricated with pads accessible by a probe station will display pad, probe, cable, and other stray capacitances that can be orders of magnitude larger than the device capacitance of interest. A more difficult problem is the actual measurement of small capacitance. A commonly available, high-performance LCR meter, like the Agilent 4284A[2], has a lower measurement range of 0.01 fF ($10^{-17}$ F) and base accuracy of 0.1 %. In practice, the lower limit of detectable capacitance, signal-to-noise, and the accuracy of a capacitance measurement depend on the measurement frequency, the test signal level, and the measurement averaging time.

We have recently begun a project to evaluate the sensitivity of available capacitance measurement instrumentation to small device capacitances and best practices to achieve optimal performance for this instrumentation. It is clear that existing off-the-shelf instrumentation will have a difficult time measuring the details of device capacitances below around 0.1 fF. Towards this end we are also looking at on-chip capacitance sensors where devices could be contacted directly to the sensor without bonding pads or probe stations. To help evaluate and compare the performance of various capacitance measurement approaches, we have developed a test chip with an array of multiple capacitors of different geometries and areas. This paper will describe the test chip structure, characterization of device capacitances, and how the chip could be used to extend capacitance metrology.

**CHIP LAYOUT AND FABRICATION**

The test chip was designed by using the design rules for the 1.6 µm complementary metal-oxide-semiconductor (CMOS) process from AMI

[1] Currently at the University of Illinois Urbana-Champaign.

[2] Certain commercial equipment, instruments, or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the materials or equipment used are necessarily the best available for the purpose.

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

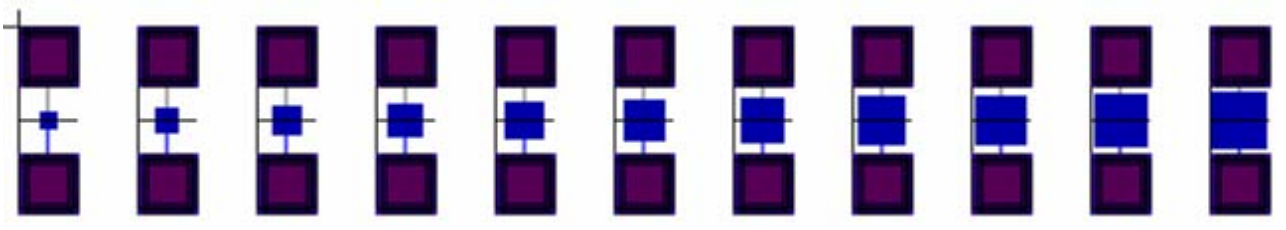


Fig. 1. A typical series of 11 devices. Bonding pads are at the top and bottom, with the MIM device in the middle.

| Series | Min (fF) | Mid (fF) | Max (fF) | Notes |
|---|---|---|---|---|
| 1 | 11.3 | 25.5 | 39.7 | Minimum size series |
| 2 | 77.9 | 99.2 | 119.7 | Variation around 100 fF |
| 3 | 159.4 | 201.9 | 250.8 | Variation around 200 fF |
| 4 | 405.6 | 497.3 | 595.7 | Variation around 500 fF |
| 5 | 794.8 | 995.9 | 1190.7 | Variation around 1000 fF |
| 6 | 987.4 | 997.3 | 1011.5 | Small variation around 1000 fF |
| 7 | 63.8 | 102.0 | 148.8 | Variation around 100 fF, square devices |
| 8 | 148.8 | 204.7 | 269.2 | Variation around 200 fF, square devices |
| 9 | 425.0 | 516.4 | 616.3 | Variation around 500 fF, square devices |
| 10 | 892.5 | 1022.8 | 1161.7 | Variation around 1000 fF, square devices |
| 11 | 11.3 | 119.7 | 1133.3 | Maximum range series |
| 12 | | 101.9 | | Variation in perimeter-to-area ratio |
| 13 | 102 | 612 | 1122 | Equal 102 fF steps |
| 14 | 102 | 102 | 102 | Repeatability at 102 fF |
| 15 | 516.4 | 516.4 | 516.4 | Repeatability at 516 fF |
| 16 | 1022.8 | 1022.8 | 1022.8 | Repeatability at 1022 fF |

Table 1. Designed oxide capacitance of each series of MOS capacitors.

Semiconductor available from the MOSIS Foundry Service. MOS capacitors were chosen for their voltage dependence and metal-insulator-metal (MIM) capacitors were chosen for their smaller total capacitance values. Both types of devices were designed with probe pads to each side of the capacitor. MOS devices were encircled by a highly doped guard band to prevent depletion edge spreading. Devices were arranged to facilitate automatic probing.

Sixteen series of 11 MOS devices were fabricated as summarized in Table 1. Series 4 of the MIM devices is shown as Figure 1. The ability to resolve a given capacitance value was tested by producing a range of devices which varied systematically around the target oxide capacitance, $C_{ox}$, according to the following sequence: 80 %, 90 %, 95 %, 98 %, 99 %, 100 %, 101 %, 102 %, 105 %, 110 % and 120 % of $C_{ox}$. Similar MIM devices were fabricated by using the intermetal dielectric, resulting in correspondingly smaller capacitances. This design yields a total of 176 devices. Two groups (A and B) of MOS and MIM devices were laid out on each chip.

## MEASUREMENT RESULTS

The MOS devices were first measured with an Agilent 4284A LCR meter. Capacitance-voltage curves were measured over a range of frequencies and test signals to determine optimum measurement conditions. C-V curves measured at 100 kHz with a 50 mV test signal and dc bias swept from -3 V to +3 V in 50 mV steps were found to have minimal series resistance effects and to produce stable capacitance values from relatively high speed measurements. Oxide capacitance was estimated by averaging the measured capacitance over a small range of voltages in accumulation. Data were acquired using an automatic probe station.

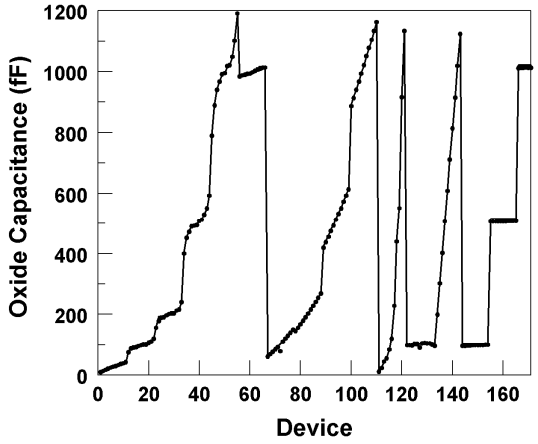



Fig. 2. Calculated oxide capacitance (solid line) and measured maximum capacitance (points) versus device number.

Excellent agreement between the designed and measured MOS oxide capacitances was seen. $C_{ox}$ was close to the value reported from the MOSIS process monitor. Figure 2 shows the measured and modeled $C_{ox}$ values for one complete set of devices. For a sample of ten different sets of capacitors, the slope of the measured and designed $C_{ox}$ was 0.96 with an intercept of 137 fF. The variation of the slope from 1 is probably mostly due to how we estimated $C_{ox}$. The intercept is close to the simulated pad plus stray capacitance of 133 fF. In general, capacitance is resolved to within about 0.5 fF (500 aF). This resolution is sufficient to see variations due to the gate metal to substrate fringe capacitance (0.5 fF to 4.5 fF, depending on device perimeter).

When the voltage is biased in accumulation, the total capacitance of each MOS device is the parallel sum of the pad capacitance, $C_{pad}$, the perimeter fringe capacitance, $C_{fringe}$, and the oxide capacitance, $C_{ox}$. The test structures are designed so that the pad capacitance is MOS geometry independent, the fringe capacitance is proportional to the device perimeter, $P$, and the oxide capacitance is proportional to the device area, $A$.

$$C_{total} = C_{pad} + P \times C_{fringe} + A \times C_{ox} \qquad [1]$$

Since most of the devices are approximately square, we introduce a parameter $S$, where the area, $A$, of the device is $S^2$, the perimeter of the device is $kS$, and $k$ is the perimeter-to-$A^{½}$ ratio of the device. The $k$-values of our devices vary from 4 (for a square device) to 12.3 (for a high aspect ratio rectangle), with an average of 4.14. Dividing by A and substituting $S^2$ for A yields:

$$\frac{C_{total}}{A} = \frac{C_{pad}}{S^2} + \frac{kC_{fringe}}{S} + C_{ox}\,. \qquad [2]$$

Hence, a plot of 1/S versus the measured capacitance per unit area should yield a 2nd order polynomial curve with a y-intercept of $C_{ox}$, a slope of $C_{fringe}$, and a curvature of $C_{pad}$. Figure 3 shows this universal capacitance curve for all the devices in a series on a single chip. For this chip we get a $C_{ox}$ of 1.10 fF/μm$^2$ (compared to 1.12 fF/μm$^2$ reported for the MOSIS process monitor), $C_{fringe}$ of 11 aF/μm (compared to 34 aF/μm for the monitor), and a $C_{pad}$ of 131.5 fF (compared to our simulated value of 133 fF). This excellent agreement confirms that our test chip can function as a reliable test vehicle for evaluating the performance of capacitance measurement circuits applied to on-chip test structures.

Chip-to-chip variation was less than ±10 % for $C_{ox}$ and $C_{pad}$, though $C_{fringe}$ was difficult to extract precisely using this technique without time consuming remeasurement, or exclusion, of outlier devices. This is probably because $C_{fringe}$ is no more than 1 % of the other capacitances. Figure 4 shows a systematic variation between the group A and group B MOS devices seen across all chips received from MOSIS. The residual capacitance is the difference in capacitances between equivalent devices in group A and in group B. For each of the 16 series of devices, a similar residual capacitance is seen and is correlated with the position of the device on the chip. Since the residual capacitance is between two identically designed devices, the observed variation must be due to some systematic process variation. These small variations make comparison of measured and designed capacitance more challenging.

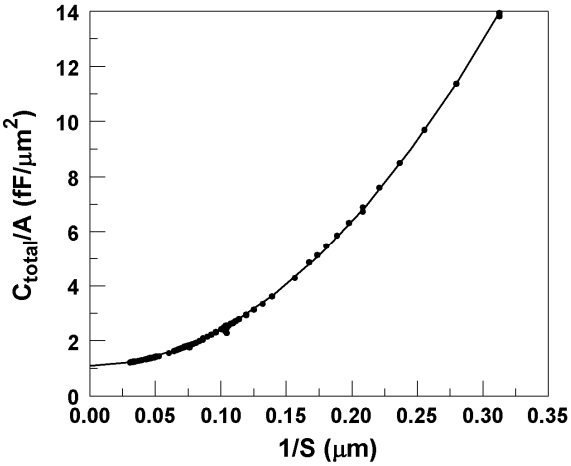



Fig. 3. Universal capacitance curve for chip C3B. Points are measured data, and the curve is the fit to data.

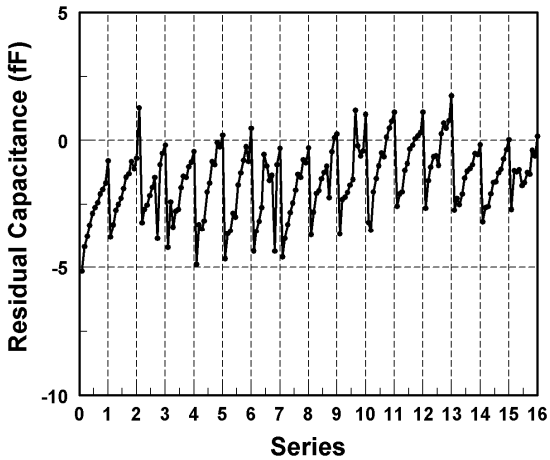



Fig. 4. Residual capacitance between devices with identical designs in different groups on the same chip. Each series consists of 11 devices.

The MIM devices display similar behavior, with corresponding smaller capacitances; their design capacitances range from 240 fF down to 0.38 fF. These devices were measured with both the Agilent 4282A LCR meter and an Andeen-Hagerling AH2700A ultra-precision capacitance bridge. In order to optimize the measurement performance of both

instruments, these devices were measured at 1 kHz, using a 1 V test signal and 5 minutes averaging time, substantially longer than required to resolve the MOS capacitance-voltage responses. Initial analysis of the data by plotting all the measured capacitances versus the design capacitance yielded lines with slopes of around 1.22 and intercepts of ~15 fF. The variation of the slope from 1 is due to a difference in the thickness of actual intermetal dielectric compared to the value used for calculating the initial model. The 15 fF intercept represents the stray probe and cable capacitance remaining after instrumentally removing the pad capacitance. Capacitances measured by the two instruments differed by about 1.2 fF over the range of devices measured. After taking this offset into account, the two instruments agreed to within an average of 50 aF. A summary of the 77 devices in the MIM series is included as Fig. 5.

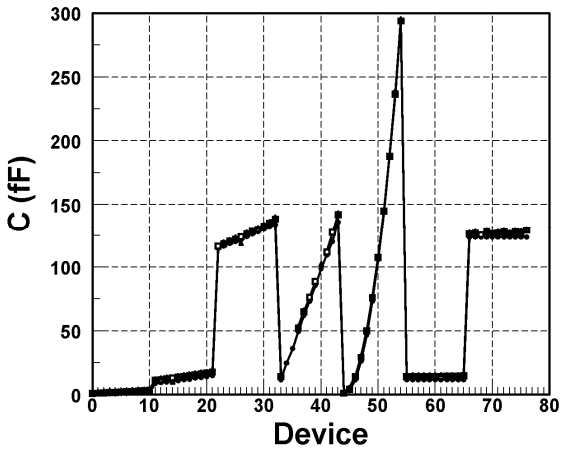



Fig. 5. Calculated and measured MIM capacitances versus device series. Theoretical data corrected for stray capacitance (solid circles) and measured capacitance for the two instruments (solid squares and open circles) are plotted.

The slope and intercept of the measured capacitance versus the design capacitance change slightly for each series of the MIM devices. This implies that as these instruments autoscale they have slightly different sensitivities to capacitance.

## SUMMARY AND FUTURE APPLICATIONS

We have designed and fabricated a test chip which enables us to measure variations in capacitance down to the 0.1 fF level. This level of performance meets our needs for a test structure to quantify the behavior of capacitance measurement instruments. This test chip will be used to evaluate various approaches to small capacitance measurement that exceed the lower bounds of the existing meters and can be implemented in a single chip or hybrid configuration. We are particularly interested in approaches that measure capacitance through measurement of the time constant of resistance-capacitance (RC) circuits or resonant frequency shifts of LCR circuits. A second version of this test chip in a more aggressive technology allowing smaller capacitances is planned.

## ACKNOWLEDGMENTS

Chong Jiang participated in this project supported by the National Science Foundation under Agreement No. EEC-0754355: "REU Site: SURFing the Electronics and Electrical Engineering Laboratory."

## REFERENCES

[1] C.-S. Tang, S.-M. Yu, H.-M. Chou, J.-W. Lee and L. Yiming, Simulation of electrical characteristics of surrounding- and omega-shaped-gate nanowire FinFETs, ***Pro. 2004 4th IEEE Conf. on Nanotech.***, pp. 281-283, (Munich, Germany, 16-19 June 2004).

# An enhanced model for thin film resistor matching

T. G. O'Dwyer, M.P. Kennedy*

Analog Devices Inc.
804 Woburn St., Wilmington, MA, U.S.A.
*Department of Microelectronic Engineering,
University College Cork
and Tyndall National Institute, Cork, Ireland.

## ABSTRACT

This paper presents an improved model which estimates the geometry required to achieve a desired matching target for rectangular resistors in a semiconductor process. A methodology is explained for estimating the model parameters involved. Measured data is presented which covers an extensive range of geometries on a particular thin film process, and the estimation methodology is followed to derive appropriate model parameters. Using this new model, insights into the underlying error sources for the process are obtained.

## INTRODUCTION

While great strides have been made in design techniques, such as sigma delta methodologies, to reduce circuit dependence on component matching, nonetheless the problem of fabricating closely-matched circuit elements remains a key challenge in microelectronics. Matching of transistors, capacitors and resistors is still a technology performance limit in applications such as Nyquist analog-to-digital converters (ADCSs), and digital-to-analog converters (DACs) for level setting applications, as well as precision gain and attenuation circuits [1]. While calibration and trimming techniques can be used to alleviate these difficulties, such methods tend to be relatively expensive compared to non-calibrated circuits in terms of silicon area, device calibration time, and the use of expensive equipment such as laser trimmers.

In this work we develop a matching model for an advanced thin film resistor process and from this derive insights into the underlying sources of error.

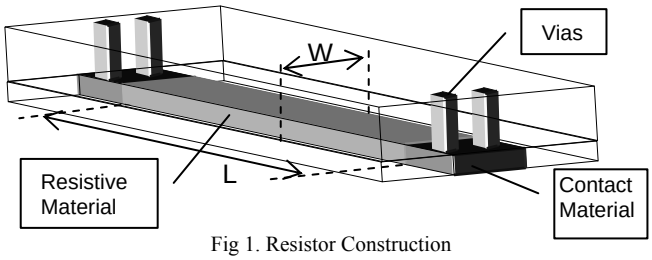



Fig 1. Resistor Construction

## THIN FILM RESISTOR PROCESS

The resistor process we examine is based on commercial thin film technology on silicon wafers [2]. This type of thin film can be used in conjunction with any geometry node of a CMOS or bipolar process in situations where precision resistors are required. The resistor structure in this particular case is fabricated with three layers, as shown in Fig. 1: (1) thin film width definition layer, (2) conductive contact definition layer, and (3) via contact definition layer.

## THIN FILM RESISTOR TEST MASK

A resistor matching test mask was generated for the process based on pairs of matched resistors connected in series, as shown in Fig. 2. Care was taken in the design of the mask to account for known topographical effects which affect matching. To reduce the effect of contact resistance, pairs of bond pads are used to implement a standard four-terminal (Kelvin) configuration at each measurement point, as shown in Fig. 2. In addition, 'dummy' resistors are employed on both sides of each pair, as shown, to ensure uniformity of resistor edge etching.

The test mask uses a broad range of twelve width and length

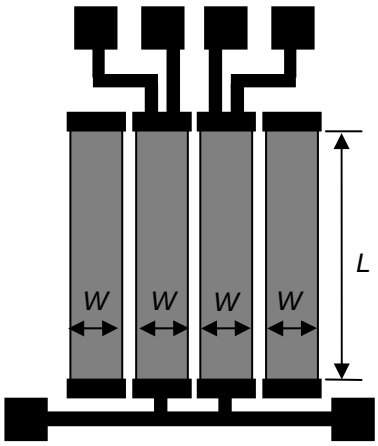



Fig 2. Example geometry test structure

combinations covering the spread of values likely to be seen in practical circuits.

## RESULTS SUMMARY

Using this test mask, the dc resistance value of each resistor was measured for each geometry on approximately three hundred test dice across ten wafers. From these, the mean resistance values and standard deviations of mismatch were calculated, where mismatch is defined as the ratio of the resistances for each pair. The results are shown in Table I. To protect proprietary information in relation to the process, the results have been normalized to the smallest geometry device i.e. 3µm x 1.9µm.

As outlined for transistors in [3] and [4], the standard deviations of the mismatches can be plotted against the reciprocal square root of the area, (i.e. $L$ times $W$), and one would expect to find a linear relationship. The results are presented graphically in Fig. 3.

As can be seen, a linear fit to the data (dashed line) gives a reasonable first order fit. However, it results in a serious error for large areas, where the best model accuracy is usually required. On the other hand, weighting the fit in favor of large areas (the solid line) results in over-optimistic estimates of mismatch in the case of small areas. Clearly, the simple linear model is inadequate in this case.

## MODEL DEVELOPMENT

We have developed an enhanced model that provides a significantly better fit to the measured data.

We begin by assuming that the resistance of a rectangular block of material of uniform resistivity is given by

$$R = \rho\,[L + \Delta L] / [W + \Delta W], \qquad (1)$$

where $R$ is resistance (ohms), $\rho$ is the resistivity per unit square (ohms/square), and $L$ and $W$ are the drawn length and width, respectively.

TABLE I
RESULTS SUMMARY

| Length, $L$ (µm) | Width, $W$ (µm) | Mean (Ohms) | Standard Deviation |
|---|---|---|---|
| 3 | 1.9 | 1.0000 | 1.0000 |
| 5 | 1.9 | 1.5851 | 0.6624 |
| 10 | 1.9 | 3.0564 | 0.3549 |
| 25 | 1.9 | 7.4738 | 0.1640 |
| 50 | 1.9 | 14.825 | 0.0878 |
| 100 | 1.9 | 29.533 | 0.0567 |
| 200 | 1.9 | 59.029 | 0.0376 |
| 200 | 3 | 37.637 | 0.0273 |
| 200 | 5 | 22.695 | 0.0257 |
| 200 | 10 | 11.375 | 0.0235 |
| 200 | 25 | 4.5515 | 0.0169 |
| 200 | 50 | 2.2761 | 0.0200 |

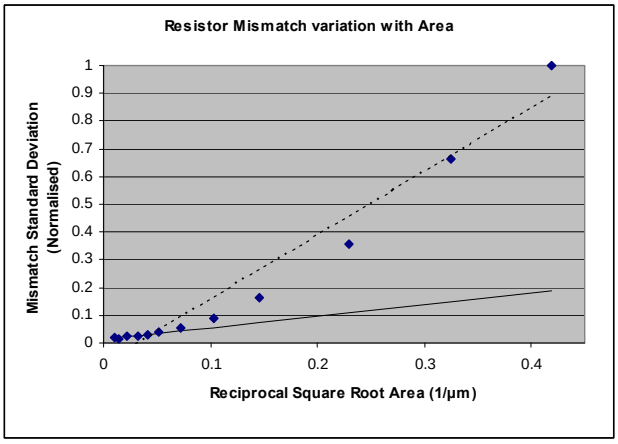



Fig 3. Resistor Mismatch variation with area

In (1), $\Delta L$ and $\Delta W$ are systematic errors in the length and width, respectively, between the drawn geometry and the physical geometry on silicon.

In practice, all three terms in (1), i.e. $\rho$, $[L + \Delta L]$, and $[W + \Delta W]$, are subject to variation. Minor perturbations in film thickness and uniformity will result in changes in $\rho$, while oxide thickness and etching variations modulate the length and width terms. In developing the model, we assume that all three perturbations can be considered as independent, normally-distributed random variables, with respective means: $\mu_\rho$, $\mu_L$, $\mu_W$, and standard deviations: $\sigma_\rho$, $\sigma_L$, $\sigma_W$. Based on this assumption, the resistance is a function of three normal random variables. The resulting probability density function is no longer a simple normal distribution [5]. However, if $\sigma_W$ is small compared to $[W + \Delta W]$, then a simplified expression for its mean and standard deviation can be derived.

In particular,

$$\mu_R = \mu_\rho\,[\mu_L / \mu_W], \qquad (2)$$

where $\mu_R$ is the resulting mean resistance. The standard deviation of the resistance, $R$, can be approximated as follows

$$[\sigma_R/\mu_R]^2 = [\sigma_L/\mu_L]^2 + [\sigma_W/\mu_W]^2 + [\sigma_\rho/\mu_\rho]^2, \qquad (3)$$

where $\sigma_R$ is the standard deviation of the resistance. Note that, by definition of $\Delta L$ and $\Delta W$, the following relations apply:

$$\mu_L = [L + \Delta L] \text{ and } \mu_W = [W + \Delta W]. \qquad (4)$$

Clearly, the first two terms in (3) are geometry dependent. The third term is also geometry dependent by the following reasoning. Consider a resistor of precise length, $\mu_L$, and width, $\mu_W$. This can hypothetically be divided into an array of arbitrarily small hexahedra, each with a base area of one unit square, and a height equal to the thickness of the film. Each hexahedron will have small variations in film thickness and uniformity corresponding to those of the overall resistor at that location, and will therefore have a particular resistivity value associated with it. Thus, the hexahedra taken collectively across the population can be considered as a set of random variables. By the law of large numbers, these can

be considered as normal random variables. Hence, the third term in (3) can be rewritten as

$$[\sigma_\rho/\mu_\rho]^2 = [\sigma_{\rho u}/\mu_\rho]^2 / \ [\mu_L \ \mu_W], \qquad (5)$$

where $\sigma_{\rho u}$ is the standard deviation of the resistivity of a unit square of the film. Note that the term, $\sigma_{\rho u}/\mu_\rho$, is a measure of the fractional variation in resistivity and is thus an overall geometry-independent figure-of-merit defining the uniformity of the process. Substituting (4) and (5) into (3) results in the following two relationships:

$$\mu_R = \mu_\rho \, [L + \Delta L] \, / \, [W + \Delta W], \qquad (6)$$

and

$$[\sigma_{Rm}]^2 = 2[\sigma_L/(L+ \Delta L)]^2 + 2[\sigma_W/(W + \Delta W)]^2 + 2 \, (\sigma_{\rho u}/\mu_\rho)^2/[(L + \Delta L)(W + \Delta W)], \qquad (7)$$

where $\sigma_{Rm}$ is the fractional standard deviation of the mismatch between the resistor pairs. The factor 2 in (7) arises from the fact that we are interested in the standard deviation of the resistor mismatch, rather than its absolute value.

Thus, we have derived a mismatch model (7) for rectangular resistors which depends on six process parameters, as well as the chosen geometry parameters, *L* and *W*.

Note that (7) contains three elements: (a) a term related solely to the quality of the length etching, (b) a term related solely to the width etching, and (c) a term related to the uniformity of the film resistivity. Note also that this latter term is inversely proportional to area. Therefore, (7) can be considered as an enhanced version of the linear Pelgrom model for random mismatch [4]. In effect, the linear model [4] only accounts for variations in film uniformity, while the model in (7) also takes account of the effects of etching on the length and width edges.

The model also has similarities to one derived by a propagation of variance approach [6]. However, (7) is an improvement on [6] since it retains the systematic length and width error terms, $\Delta L$ and $\Delta W$. It is important to include these two parameters in matching models because ignoring them can lead to significant errors in the estimation of the standard deviation parameters. This arises from the fact that, in order to stretch the process to its limits, typical resistor test structures usually contain several geometries which have minimum length and width. Such structures have maximum errors caused by $\Delta L$ and $\Delta W$, and therefore data from these geometries can lead to significant errors in model parameter estimation.

## PARAMETER ESTIMATION PROCEDURE

While a number of methods exist for model parameter estimation based on a data set, we have chosen a linear approximation method for ease of implementation. This is based on the observation that error perturbations due to process variation are likely to be small, and thus we can make use of Taylor series approximations.

In (6), the denominator can be expanded as a Taylor series. Multiplying by the numerator and ignoring higher order error terms results in the following approximation:

$$\mu_R = (\mu_\rho)[L /W] + (\mu_\rho \, \Delta L) \, [1/W] \ + (\mu_\rho \Delta W)[-L /W^2]. \qquad (8)$$

Clearly, if the mean resistance values are measured for a variety of appropriate *W* and *L* geometries, one can generate an over-determined set of linear equations, and from this the three unknowns, $\mu_\rho$, $\Delta L$ and $\Delta W$ can be found by a least squares surface fit. Thus, we have a method to estimate the systematic length and width errors, as well as the average film resistivity.

Equation (7) can be rewritten as follows:

$$(\sigma_{Rm}^2)/2 = (\sigma_L)^2 \, [1/(L+ \Delta L)^2] + (\sigma_W)^2 \, [1/(W + \Delta W)^2] + (\sigma_{\rho u})^2 \, [1/\{\mu_\rho^2 \, (L + \Delta L)(W + \Delta W)\}]. \qquad (9)$$

If the standard deviations of matching are known for the same set of *W* and *L* geometries, then by substituting the values of $\mu_\rho$, $\Delta L$ and $\Delta W$ found in the previous step, one can again generate an over-determined set of linear equations, and the remaining three parameters, $\sigma_L$, $\sigma_W$, and $\sigma_{\rho u,}$ can be estimated.

## FITTING THE MODEL

We applied the procedure described above to the data set summarized in Table I and Fig. 3. The estimated $\Delta L$ and $\Delta W$ terms turned out to be negligible compared to practical geometries in use, and the $\sigma_W$ term was also very small.

For the normalized data of Table I, the simplified model is as shown in (10), where *L* and *W* are in micrometers:

$$\sigma_{Rm} = \sqrt{\frac{1.02\text{x}10^{-3}}{L^2} + \frac{4.58\text{x}10^{-5}}{LW}} \ \text{x}100 \ \ \%. \qquad (10)$$

The fit to the data resulting from this model is shown graphically in Fig. 4. Note that the last data point on the right of the graph, which corresponds to the 3µm x 1.9µm geometry in Table I, was not included in the least squares

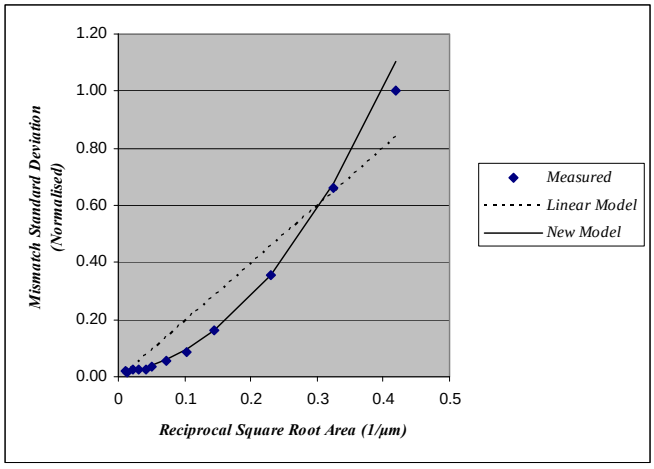



Fig.4. Model performance versus measured data

fitting as its geometry is outside the normal design rules for this process.

As can be seen from Fig. 4, the new model (10) provides a close match to the data over the full spread of geometries under investigation, and represents a considerable improvement over the simple linear relationship [4].

From the foregoing, we conclude that, for this particular process and range of geometries, the systematic length and width errors (i.e. $\Delta L$ and $\Delta W$ resp.) can be ignored when estimating mismatch. Similarly, the random variation in the width etching can also be neglected, since $\sigma_W$ is negligible for the practical geometries under consideration. We conclude that the two most significant contributors to the standard deviation of the mismatch error are (a) the statistical variation in the length etching, $\sigma_L$, and (b) the fractional variation in the film thickness or uniformity, $\sigma_{\rho u}/\mu_\rho$.

Since the model parameters are measures of physical properties of the process in each case, our methodology provides a powerful tool for identifying potential strategies for improving the process. To validate a proposed improvement, the procedure described can be applied and the effect of the process change on the length etching, width etching, resistivity uniformity, etc. can be isolated and quantified.

## CONDITION FOR LENGTH TERM DOMINANCE

Since (10) consists of a length etch term related solely to length, and a film uniformity, or "Pelgrom", term related to area, it is instructive to ask: under what conditions does one dominate over the other?

For the second term to dominate in this example, we require that

$$4.58\text{x}10^{-5}/LW > 1.02\text{x}10^{-3}/L^2, \qquad (11)$$

or equivalently, that

$$L/W > 22 \text{ squares}. \qquad (12)$$

Thus, the dominance of one term over the other simply depends on the target resistance value. For resistances significantly beyond that defined in (12), the length term can be neglected, and (10) reduces to the conventional Pelgrom relationship. Conversely, if the resistance is significantly less than shown in (12), then the Pelgrom term can be neglected and (10) reduces to a linear relationship with the reciprocal of length. Note that under this latter condition, the width has no effect on matching, which runs contrary to what one would expect with the simple Pelgrom model.

The model (10) is in marked contrast to the results reported for polysilicon resistors in [7]. In that case, the authors report a linear relationship between the standard deviation of the mismatch and the reciprocal square root of the area, as predicted by the conventional Pelgrom model [4]. In addition, they demonstrate a correlation between mismatch and polysilicon grain size. In our work, the reciprocal square root of area term is less dominant for the thin film resistors studied, suggesting that the proprietary film has a more uniform structure than polysilicon.

## GRAPHICAL REPRESENTATION

Useful insight for circuit designers can be gained by examining (10) from a graphical viewpoint. Fig. 5 shows the locus of lengths and widths (in normalized units) derived from (10) which are required to achieve a particular target value for the standard deviation of the mismatch, in this case 0.1%. Note that, as the width is increased beyond approximately 0.6 units, the length required to achieve the target matching remains almost constant and is largely independent of width. This corresponds to the region where the width has little effect on matching, and where the Pelgrom term becomes less significant, as described above.

To achieve the required mismatch target, any combination of $W$ and $L$ above the locus will suffice. However, in practice, there is usually the additional requirement of achieving a particular resistance value. In Fig. 5, a positive resistance appears as a line with positive slope and with its intercept at the origin. Any combination of $W$ and $L$ along the constant resistance line inside the solution region will suffice to meet both constraints.

Finally, it is usually desirable to minimize the area of the resistor for cost reasons. Therefore, the optimum combination of $W$ and $L$ is located at the intersection of the constant resistance line and the required maximum mismatch locus.

In general, only a limited number of target mismatches are required, for example, 0.4% (8 bit), 0.2% (9 bit), 0.1% (10 bit), etc. In this case, a set of contours for a particular process can be plotted on a single graph, as shown in Fig. 6. The geometry required for a particular set of matching and resistance constraints can simply be read from the resulting graph.

Alternatively, an analytical method can be used. Equation (10) and the resistor equation, $R = L/W$ squares, can be solved as simultaneous equations to express the required values of $W$ and $L$ directly in terms of the target matching and the

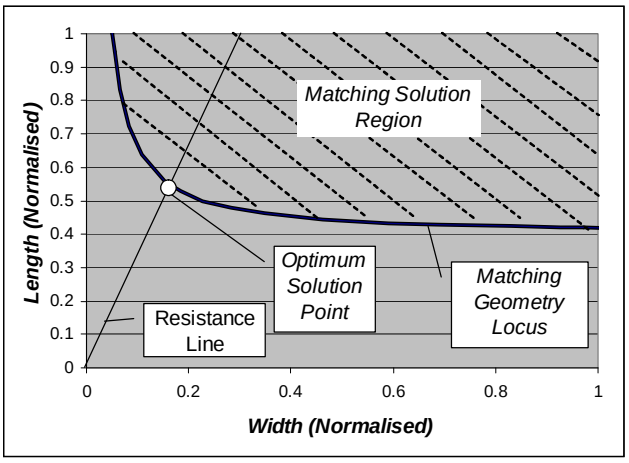



Fig. 5. Geometry locus for 0.1% mismatch

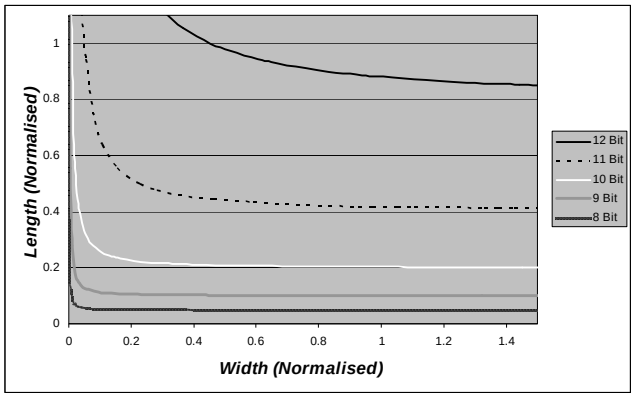



Fig. 6. Process mismatch loci for 8 to 12 bit matching

required resistance. In this case, the resulting expressions are as follows:

$$L = (100/\sigma_{Rm}) \sqrt{(1.02\text{x}10^{-3} + 4.58\text{x}10^{-5}R)}\ \mu\text{m}, \quad (13)$$

$$W = L/R\ \ \mu\text{m}, \quad (14)$$

where $R$ is the desired resistance in squares, and $\sigma_{Rm}$ is the desired maximum standard deviation of mismatch in percent.

## CONCLUSION

We have presented an enhanced matching model for rectangular resistors and have defined a linear approximation procedure which permits easy estimation of its parameters. We have demonstrated an example of the resulting model and parameter estimation methodology for a thin film resistor process. The model parameters relate to specific physical attributes of the process, so the technique can be used to drive improvements in process performance. The resulting model fit, which has terms that depend on $1/L^2$, as well as $1/WL$, yields better agreement with measured data than the simple area model described in [4]. By examining the relative dominance of its length and area terms, we have presented insights into the main causes of mismatch for the process. In addition, we have developed both a graphical and an analytical method to ease the problem of optimum resistor design.

## ACKNOWLEDGEMENT

The authors wish to express their gratitude to Mr. Frank Poucher, Advantec Ltd., and his team for provision of the test mask and data used as the basis of the model development.

## REFERENCES

[1] P.R. Gray, R. G. Meyer, "Analysis and design of analog integrated circuit", John Wiley & Sons, New York, 3rd Ed., ISBN 0-471-57495-3, pp. 451–453, 1993.

[2] D. Smith, "Thin-film Deposition: Principles and Practice", ISBN 0-07-058502-4, McGraw-Hill, 1995.

[3] M.J.M. Pelgrom, A.C.J. Duinmaijer, A.P.G. Welbers, "Matching properties of MOS transistors," IEEE Journal of Solid-State Circuits, Vol. 24, Issue 5, pp.1433–1439, Oct 1989.

[4] M.J.M. Pelgrom, H.P. Tuinhout, M. Vertregt, "Transistor matching in analog CMOS applications," International Electron Devices Meeting, 1998, IEDM '98 Technical Digest, pp. 915–918, 6-9 Dec. 1998.

[5] A. Papoulis, "Probability, random variables and stochastic processes", ISBN 07-48448-1, McGraw-Hill, pp.196-8 & p.205, 1965.

[6] P. Drennan, "Diffused resistor mismatch modeling and characterization", IEEE Bipolar/BiCMOS Circuits and Technology Meeting, pp. 27-30, 26-28 Sept. 1999.

[7] R. Thewes, R Brederlow, C. Dahl, U. Kollmer, C. Linnenbank, B. Holzapfl, J. Becker, J. Kissing, S. Kessel, W. Weber, "Explanation and quantitative model for the matching behavior of poly-silicon resistors", International Electron Devices Meeting, pp.771-774, 6-9 Dec. 1998.

[8] P. Kinget, M. Steyaert, "Impact of transistor mismatch on the speed-accuracy trade-off of analog CMOS circuits" Custom Integrated Circuits Conference, 1996., Proceedings of the IEEE 1996, pp.333 – 336, 5-8 May 1996.

[9] M. Steyaert, V. Peluso, J. Bastos, P. Kinget, W. Sansen, "Custom analog low power design: the problem of low voltage and mismatch", Custom Integrated Circuits Conference, 1997., Proceedings of the IEEE 1997, pp.285 – 292, 5-8 May 1997.

[10] J. Oehm, U. Grunebaum, K. Schumacher, "Overcoming insufficiencies of first order mismatch modeling," ESSCIRC '98. Proceedings of the 24th European Solid-State Circuits Conference, pp. 508–511, 22-24 Sept. 1998.

# Application of Matching Structures to Identify the Source of Systematic Dimensional Offsets in GHOST Proximity Corrected Photomasks

S. Smith*, A. Tsiamis*, M. McCallum†, A.C. Hourd‡, J.T.M. Stevenson* and A.J. Walton*

* Institute for Integrated Micro and Nano Systems, School of Engineering,
Scottish Microelectronics Centre, The University of Edinburgh, EH9 3JF, UK
Email: Stewart.Smith@ed.ac.uk

†Nikon Precision Europe GmbH, Appleton Place, Appleton Parkway, Livingston, West Lothian, EH54 7EZ, UK

‡Compugraphics International Ltd., Eastfield Industrial Estate, Glenrothes, Fife, KY7 4NT, UK

## ABSTRACT

The effects of the GHOST proximity correction process on chrome-on-quartz photomasks can prove difficult to quantify and so they are not routinely characterised. This paper presents a methodology for addressing this issue using electrical test structures designed to measure dimensional mismatch. In the past these have been used successfully to characterise standard GHOSTed photomasks, which displayed systematic offsets that were not seen on an unGHOSTed mask using the same design. In order to investigate this further, a second mask was fabricated using a variation of the GHOST process which increased the resolution of the secondary exposure to be the same as the primary pattern. This enabled the source of the previously observed systematic offset to be determined as test structures on the new mask did not show the same overall dimensional bias. However, the range of mismatch in some of the structures was increased as a result of the new process.

## INTRODUCTION

The work presented in this paper is based upon previously reported on-mask, electrical test structures for the characterisation of photomasks. Publications have included studies of advanced phase-shifting masks [1], investigations into different process proximity correction schemes [2], [3] and comparisons of different metrology techniques [4]. This work has demonstrated that direct electrical measurements of test structures on photomasks compare very well with traditional mask metrology techniques.

This paper presents new results from masks printed with an array of dimensional mismatch test structures similar to those described in [5] and [6]. These masks have been fabricated using a technique known as GHOST [7] which attempts to correct for proximity induced dimensional errors in electron-beam lithography. In this process the main e-beam exposure is followed by a secondary exposure with the inverse of the main pattern at a lower dose. This serves to equalise the background dose caused by electron scattering and reduces adverse proximity effects such as dimensional bias between isolated and dense features.

Previous mismatch results from a GHOSTed mask showed significant, systematic offsets which were ascribed to the fact that the secondary GHOST pattern was written with a larger grid size to speed up the process. In order to properly identify the source of the observed dimensional offsets, a new mask has been fabricated using a "Slow GHOST" process where the grid size is kept the same for both exposures.

## TEST STRUCTURES AND MEASUREMENTS

The mismatch test structures used in this research are pairs of Kelvin bridge resistors as shown in figure 1. The bridge sections are 600μm long with a nominal width of 0.5μm and a separation of 30μm between the pairs of resistors. Each test structure block has two sets of resistors, one arranged horizontally and one vertically.

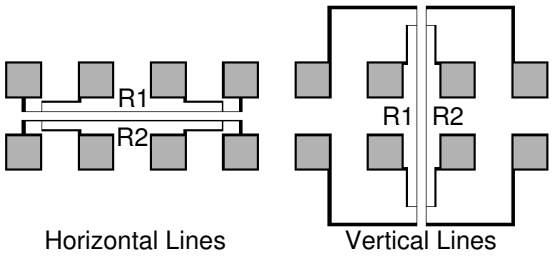



Fig. 1. Layout of mismatch bridge resistor test structures

The bridge resistors are measured by passing a current ($I$) between the outer pads and measuring the resulting voltage ($V$) with the other two pads. The resistance ($R = V/I$) can then be used to estimate the electrical width of the feature. As there are no structures on the masks which can be used to measure the sheet resistance of the chrome layer, a nominal value derived from other test masks is used. Although a cross-mask variation of $R_S$ of about 1% has been noted in the past the short range variation is significantly lower and the assumption that the sheet resistance of closely spaced lines is the same is a fair one. A nominal sheet resistance of 22.5$\Omega/\square$ has been used to obtain the electrical linewidth results in this paper. For each pair of lines the width of

resistor R2 is subtracted from that of R1 to obtain the width offset $\Delta W$. This is divided by the average width of the pair to give the mismatch figure $\Delta W/W$ (%). As this is a relative value it is unaffected by the value of $R_S$ chosen for the linewidth calculation.

Three masks were fabricated using the design shown in figure 2, which consists of an array of 357 blocks of mismatch test structures in a regular grid with a pitch of 6mm. The first of these masks (MSN7520) had no GHOST processing [5] while the second (MSN7553) used the standard GHOST process. Finally, the third mask (MSN7864) used "Slow GHOST" in order to investigate the systematic linewidth bias effects seen on previous GHOSTed masks.

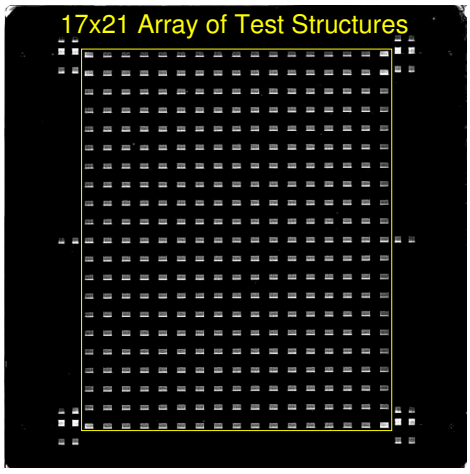
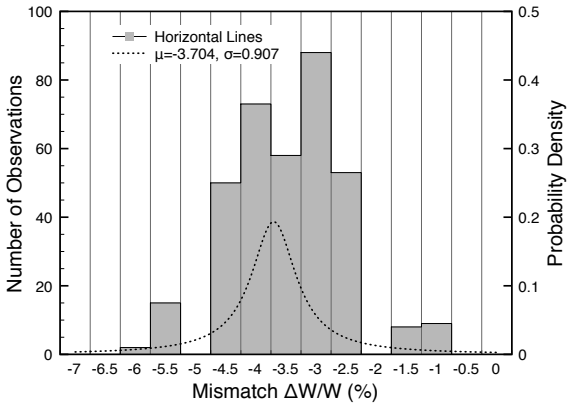
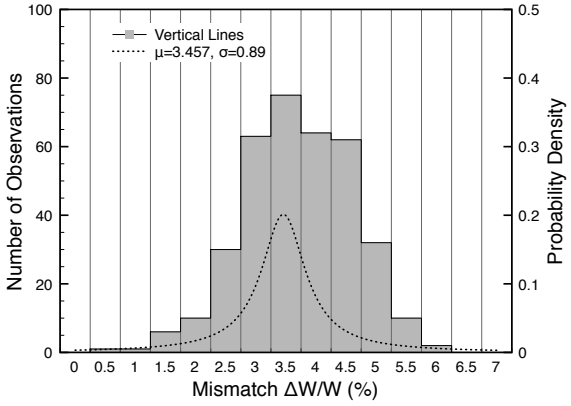



Fig. 2. Scanned image of chrome-on-quartz mismatch test mask




(a) Horizontal Lines




(b) Vertical Lines

Fig. 3. Histograms of CD mismatch results for MSN7553

## RESULTS

The full sets of vertical and horizontal mismatch structures have been measured on both MSN7553 and MSN7864 and the results are presented as histograms in figures 3 and 4 respectively. These graphs also show normal distributions fitted to each set of data where "μ" is the mean value and "σ" is the standard deviation. However, this does not indicate that the date is normally distributed.

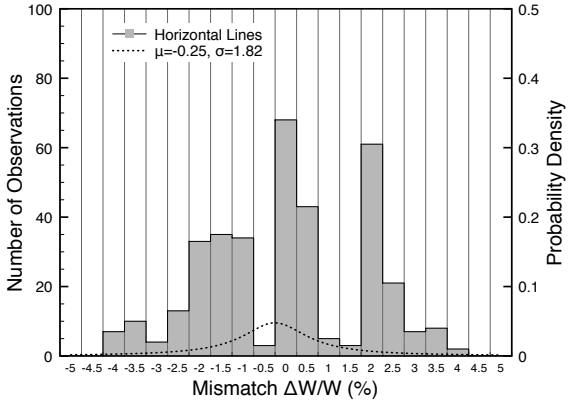
It is clear that the systematic offsets observed on previous GHOSTed masks are also seen on the mask using the standard GHOST process (MSN7553) while the results from the "Slow GHOST" mask (MSN7864) have mean values closer to zero. Table I summarises these results and includes data from the noGHOST mask (MSN7520).

The Slow GHOST process reduces the mean mismatch to a similar level to that seen on the noGHOST photomask. However, it also appears to greatly increase the range of the results from the horizontal lines. Although the histogram plots have been fitted with normal distributions it is clear that the results from the horizontal lines do not show a random, statistical variation. Therefore, the standard deviations given in figures 3(a) and 4(a) do not provide useful information. It is for that reason that the range value, representing the difference between maximum and minimum mismatch, is provided in table I. The results presented in [5] for horizontal structures on MSN7520 (noGHOST) plotted against the vertical position on the mask show good agreement between structures in the same row. Similar results have been obtained for the two GHOSTed masks and can be seen in figure 5.

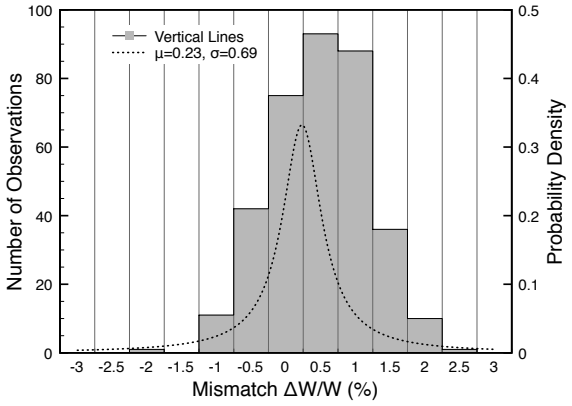
These results show a similar pattern to that observed for MSN 7520, in particular with the maximum mismatch occurring in row 4 of the mask and the minimum in row 9. A possible reason for this, suggested in [5] was the fact that there is an e-beam pattern boundary which runs between the two horizontal lines on these rows. However, there are also similar pattern boundary conditions on rows 1 and 17 which do not lead to large mismatches. The similarity between the results from the three mask for horizontal mismatch structures can be seen more clearly in figure 6 which plots the average mismatch for each row. The error bars, where visible, give the standard deviation of the mismatch across the 17 sets of structures in each row.

TABLE I
COMPARISON OF MISMATCH RESULTS FROM THREE DIFFERENT MASKS

| Mask MSN | Horizontal $\Delta W/W(\%)$ | | | Vertical $\Delta W/W(\%)$ | | |
|---|---|---|---|---|---|---|
| | Mean | σ | Range | Mean | σ | Range |
| 7520 (No GHOST) | -0.37 | 0.58 | 3.01 | 0.1 | 0.79 | 4.48 |
| 7553 (GHOST) | -3.7 | 0.91 | 4.77 | 3.46 | 0.89 | 5.05 |
| 7864 (Slow GHOST) | -0.25 | 1.82 | 7.96 | 0.23 | 0.69 | 4.19 |

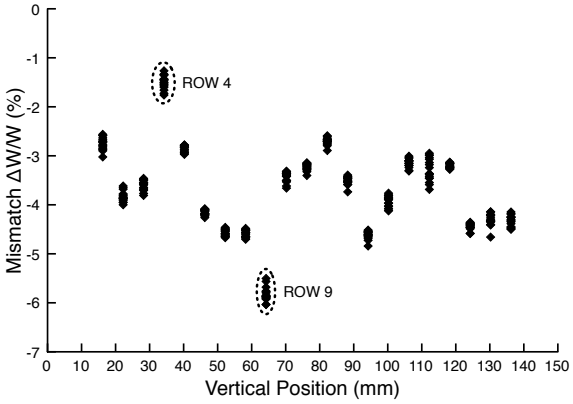



(a) Horizontal Lines

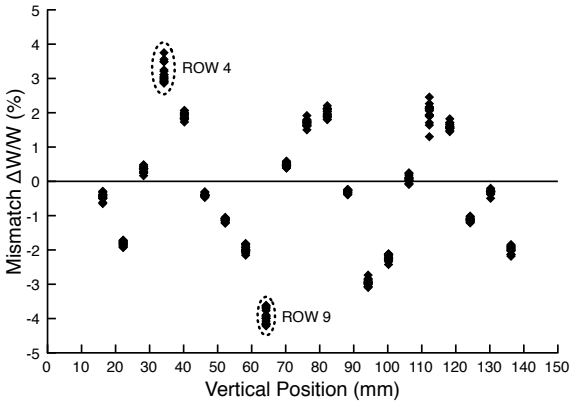



(b) Vertical Lines

Fig. 4. Histograms of CD mismatch results for MSN7864




(a) MSN7553 (GHOST)




(b) MSN7864 (SlowGHOST)

Fig. 5. Mismatch for horizontal lines plotted against vertical position

The data in figure 6 has been normalised in each case by subtracting the mean mismatch figure for the whole mask and it is clear that all three masks show the same overall pattern. However, the Slow GHOST mask, MSN7864, shows much greater variation. One possible reason for this observed effect is interaction between the primary and secondary exposures in the GHOST process. These results confirm the suspicion that the systematic offset of around ±3.5% observed in the standard GHOST mask (MSN7553) results are a result of the fact that the secondary GHOST exposure is performed at a lower resolution. This means that the GHOST pattern overlaps the main pattern by 10nm on one side of each of the bridge resistors. It is still unclear how this leads to a systematic offset between the resistors in a pair as both should have similar secondary exposure patterns. In the "Slow GHOST" mask the resolution of the secondary pattern is the same as the main pattern but the increased write time means that offsets between the two patterns are more likely. This obviously has a significant, adverse effect on the horizontal lines, increasing the mismatch range to more than ±3%, equivalent to linewidth offsets of more than ±15nm.

The vertical lines on MSN7864 do show the elimination of the systematic offset when compared with MSN7553

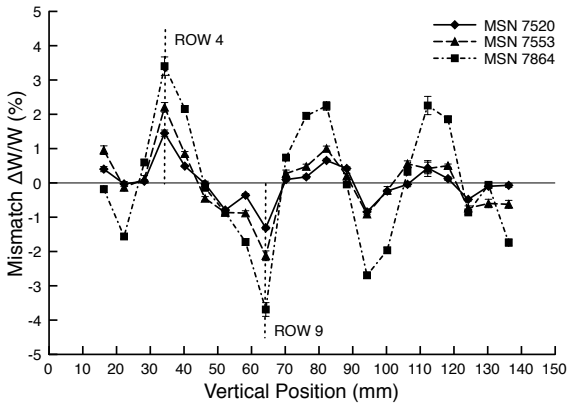
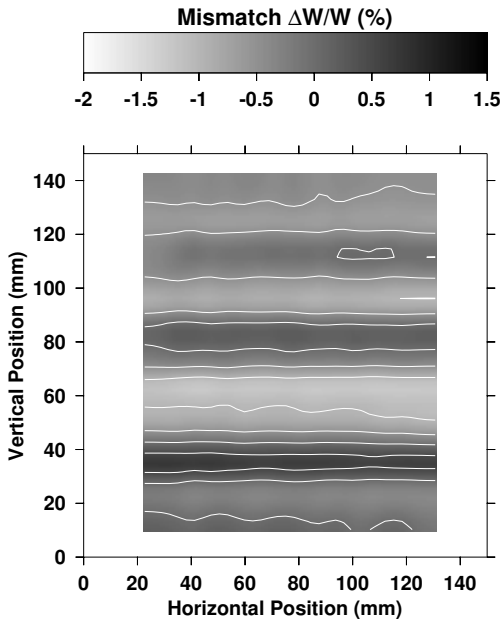



Fig. 6. Comparison of mismatch figures for horizontal structures from three different photomasks

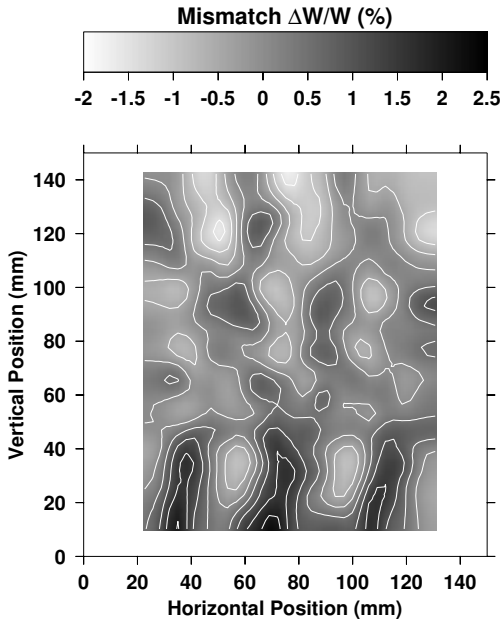
but do not show an increase in the range of the mismatch results. As was noted in [5] the pattern is split into "write segments" which are around 160μm high. This means that the vertical lines are split between several write segments while the horizontal lines are contained within a single segment. The result of this may be an averaging of adverse effects associated with the high resolution GHOST pattern. Although the mismatch results from the vertical lines appear to be randomly distributed in figures 3(b) and 4(b) this is not actually the case as can be seen in the contour plots of the mismatch from the vertical structures (figures 7(b), 8(b) and 9(b)). Similarly, clear patterns of horizontal bands can be seen in contour plots of results from the horizontal lines (figures 7(a), 8(a) and 9(a)).

## CONCLUSIONS

Electrical measurement results from mismatch resistor structures fabricated on standard chrome-on-quartz photomask plates have been presented. The masks have nominally identical patterns but have been processed differently. The first of these, MSN7520 (No GHOST), was processed without any correction for process or pattern dependent biases and full results from this mask were presented in a previous publication [5]. The second, MSN7553 (GHOST) has been prepared using a standard GHOST process commonly used to remove pattern dependent biases in electron-beam lithography. Finally, MSN7864 (Slow GHOST) used an adaption of this process in an attempt to identify the source of systematic dimensional offsets observed in previous results from mismatch test structures. It was thought that the systematic mismatch observed in GHOSTed structures was due to the fact that the secondary GHOST pattern was printed with a lower resolution than the primary pattern. The "Slow GHOST" process would use the same resolution for the secondary exposure, increasing the write time but removing any systematic offset between the patterns.



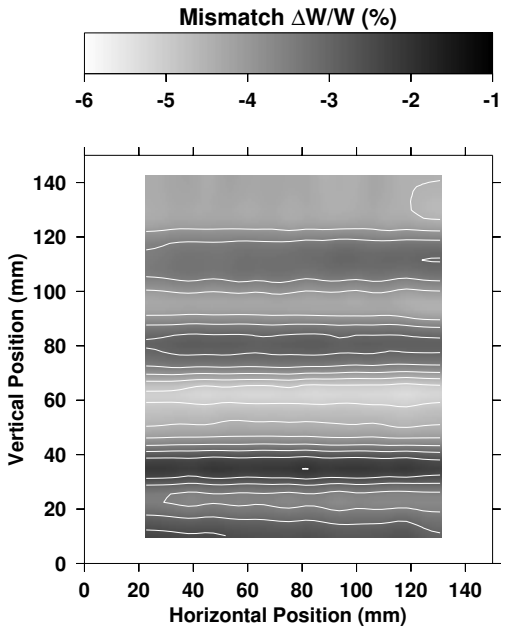

(a) Horizontal Lines



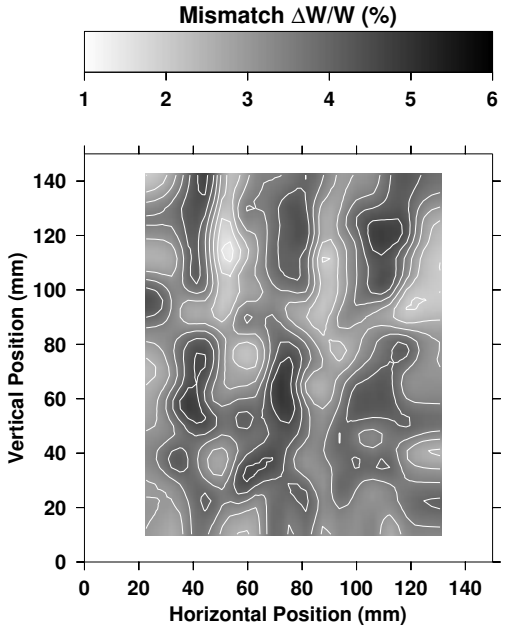

(b) Vertical Lines

Fig. 7. Contour plots of mismatch data from MSN 7520 (No GHOST).

The effect of the Slow GHOST process on both horizontal and vertical lines was to remove the overall systematic bias as expected. However, it also had the effect of significantly increasing the overall range of the mismatch results for horizontal lines. It is likely that the increased write times involved in the Slow GHOST process exacerbate the tool

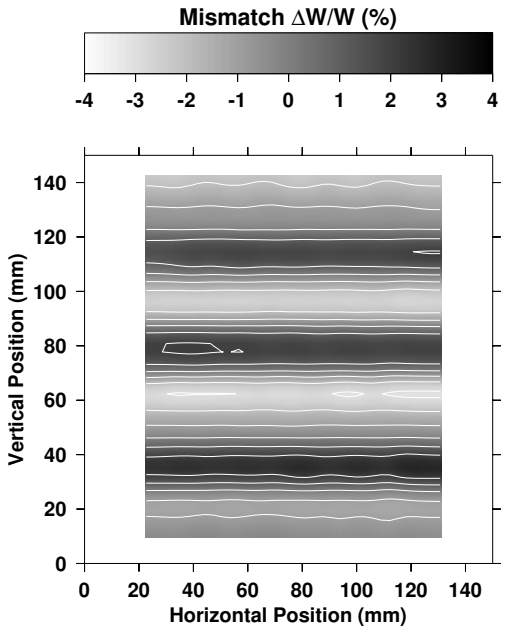



(a) Horizontal Lines

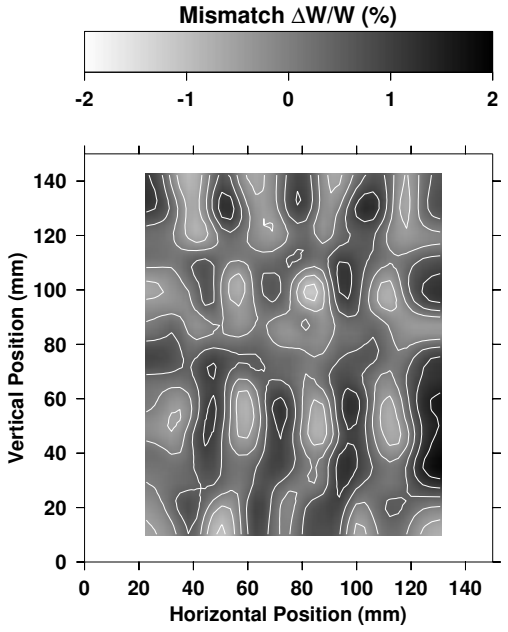



(b) Vertical Lines

Fig. 8. Contour plots of mismatch data from MSN 7553 (GHOST).




(a) Horizontal Lines




(b) Vertical Lines

Fig. 9. Contour plots of mismatch data from MSN 7864 (Slow GHOST).

induced pattern seen in figure 6. Contour plots of mismatch data from the three masks discussed in this paper show a clear, row by row, pattern in the mismatch data from horizontal lines which is similar for each mask. However, the range of the mismatch variation increases from mask to mask as the write time gets progressively longer for the No GHOST, GHOST and Slow GHOST write strategies. This suggests that the increased write time could be compromising the re-registration performance of the write tool in some way, perhaps by a mechanism such as thermal drift within the mask writing chamber. Although the mismatch data from the vertical lines appears at first to be more randomly

distributed, a clear pattern emerges in the contour plots. However, unlike the results for the horizontal structures the pattern is significantly different in each case. Overall, the application of either GHOST process has little effect on the range of the results from the vertical structures and only affects the mean value. GHOST processing has a much more significant effect on the horizontal structures, with the Slow GHOST process in particular leading to mismatches of as much as ±4% or ±20nm. This is understandable when one considers that for the vertical structures the variation in registration of the scanned beam will tend to contribute to the line edge roughness rather than shifting the whole edge as would be the case for the horizontal lines. This effect would explain the lower measured range for the vertical features. Vertical lines will have their width governed by the electron beam being blanked on or off rather than its scanned position and will therefore be less prone to linewidth variation from registration drifts.

The electrical test structures have demonstrated the capability of measuring tool and process induced dimensional mismatches in the nanometer range on photomask plates. This would be difficult with the optical metrology technology discussed in [4] which struggles to provide accurate CD measurements for structures with widths below 700nm.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Edinburgh Research Partnership in Engineering and Mathematics, Nikon Precision Europe GmbH and Compugraphics International Ltd.

## REFERENCES

[1] S. Smith, A. Tsiamis, M. McCallum, A. C. Hourd, J. T. M. Stevenson, and A. J. Walton, "Comparison of Optical and Electrical Techniques for Dimensional Metrology on Alternating Aperture Phase-Shifting Masks," *IEEE Transactions on Semiconductor Manufacturing*, vol. 21, no. 2, pp. 154–160, May 2008.

[2] A. Tsiamis, S. Smith, M. McCallum, A. C. Hourd, O. Toublan, J. T. M. Stevenson, and A. J. Walton, "Investigation of Electrical and Optical CD Measurement Techniques for the Characterisation of On-Mask GHOST Proximity Corrected Features," in *Proceedings of the IEEE International Conference on Microelectronic Test Structures (ICMTS08)*, Edinburgh, UK, March 2008, pp. 29–34.

[3] A. Tsiamis, S. Smith, M. McCallum, A. C. Hourd, O. Toublan, A. J. Walton, and J. T. M. Stevenson, "Electrical Test Structures for the Characterisation of Optical Proximity Correction," in *Proceedings of the SPIE, Vol. 6533: The 23rd European Mask and Lithography Conference (EMLC 2007)*, Grenoble, France, January 2007, p. 65330M.

[4] S. Smith, A. Tsiamis, M. McCallum, A. C. Hourd, J. T. M. Stevenson, A. J. Walton, R. G. Dixson, R. A. Allen, J. E. Potzick, M. W. Cresswell, and N. G. Orji, "Comparison of Measurement Techniques for Advanced Photomask Metrology," in *Proceedings of the IEEE International Conference on Microelectronic Test Structures (ICMTS08)*, Edinburgh, UK, March 2008, pp. 35–39.

[5] S. Smith, A. Tsiamis, M. McCallum, A. C. Hourd, A. J. W. J. T. M. Stevenson, and S. Enderling, "On-mask mismatch resistor structures for the characterisation of maskmaking capability," in *Proceedings of the IEEE International Conference on Microelectronic Test Structures (ICMTS08)*, Edinburgh, UK, March 2008, pp. 228–232.

[6] S. Smith, A. Tsiamis, M. McCallum, A. C. Hourd, J. T. M. Stevenson, and A. J. Walton, "Electrical Measurement of On-Mask Mismatch Resistor Structures," in *Proceedings of IEEE International Conference on Microelectronic Test Structures*, Tokyo, Japan, March 2007, pp. 3–8.

[7] G. Owen and P. Rissman, "Proximity Effect Correction for Electron Beam Lithography by Equalization of Background Dose," *Journal of Applied Physics*, vol. 54, no. 6, pp. 3573–3581, 1983.

# An Analysis of Temperature Impact on MOSFET Mismatch

S. Mennillo[1], A. Spessot[1], L. Vendrame[1], L.Bortesi[1]

[1] Numonyx, R&D – Technology Development
Via C. Olivetti 2, 20041 Agrate Brianza, Italy

## ABSTRACT

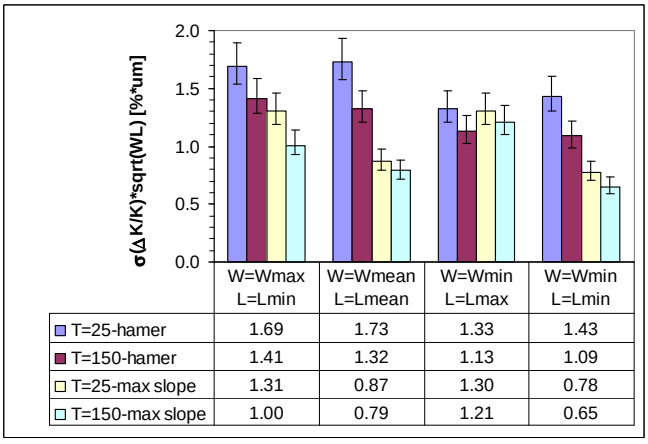
Summarizing the results collected on several technologies, we have studied the impact of temperature on MOSFET mismatch, highlighting the improvement of current gain matching properties with temperature, suggesting a possible physical explanation to this phenomenon and proposing a BSIM3 model implementation for Monte Carlo mismatch simulations.

## INTRODUCTION

Matching properties of devices have been widely studied in the last twenty years, but the impact of temperature on device mismatch has been poorly investigated up to now. Recent papers [1, 2] have proposed analytical models or simplified models for Monte Carlo mismatch simulations which attempt to include the effect of temperature on MOSFET mismatch, focusing on the implementation or on the reliability but, to authors best knowledge, no much attention has been devoted to a possible physical explanation of the observed phenomena.
Purpose of this paper is to describe the temperature impact on MOSFET mismatch, gathering the experimental results coming from different technologies, analyzing the possible root causes of current gain mismatch trend with temperature and proposing a simplified model to take account of temperature impact on mismatch when performing Monte Carlo SPICE simulations.

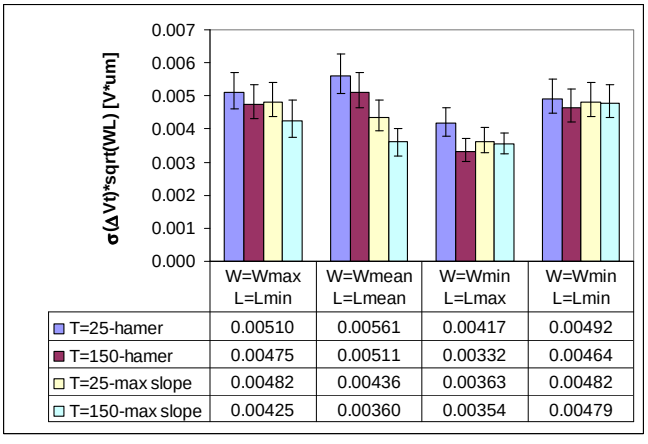
## TEST STRUCTURES AND EXPERIMENTAL EVIDENCE

We have investigated the impact of temperature on MOSFET mismatch on different technologies with oxide thickness ranging from very thin 25Å oxide to 35Å (later on called thin oxide) and up to 80Å (called thick oxide), through test structures consisting of matched pairs of different MOSFET families and geometries. The DOE for *W* and *L* covered from minimum design rules for the specific device up to tens of microns; also intermediate or mean sizes around one micron have been studied to fully cover a possible realistic design space. Matched devices were designed in the most often used common gate / common source configuration with separate drain pads, at minimum distance, with poly dummies and gate protection diode. Threshold voltage (*Vt*), current gain (*K*) and saturation current (*Id*) at several overdrives were measured at different temperatures. *Vt* and *K* were measured using both the maximum slope method and the Three Points Hamer routine [3], with the aim of verifying whether the measurement technique could have an impact on the observed mismatch when different temperatures are considered, and of avoiding misinterpretation of physical effects due to measurement artifacts (in particular for threshold and gain which are experimentally derived from manipulation of current measures at different bias points according to the different methods). Fig.1 and Fig.2 depict *K* and *Vt* mismatch, normalized w.r.t the square root of device area, for 4 devices under test.




| | W=Wmax L=Lmin | W=Wmean L=Lmean | W=Wmin L=Lmax | W=Wmin L=Lmin |
|---|---|---|---|---|
| T=25-hamer | 1.69 | 1.73 | 1.33 | 1.43 |
| T=150-hamer | 1.41 | 1.32 | 1.13 | 1.09 |
| T=25-max slope | 1.31 | 0.87 | 1.30 | 0.78 |
| T=150-max slope | 1.00 | 0.79 | 1.21 | 0.65 |

**Fig. 1**: Normalized gain mismatch, computed with Three Points Hamer routine and maximum slope algorithm, for four different 25Å oxide N-mosfet pairs. 95% confidence bars are superimposed to the plot. The number of tested dice is 106.




| | W=Wmax L=Lmin | W=Wmean L=Lmean | W=Wmin L=Lmax | W=Wmin L=Lmin |
|---|---|---|---|---|
| T=25-hamer | 0.00510 | 0.00561 | 0.00417 | 0.00492 |
| T=150-hamer | 0.00475 | 0.00511 | 0.00332 | 0.00464 |
| T=25-max slope | 0.00482 | 0.00436 | 0.00363 | 0.00482 |
| T=150-max slope | 0.00425 | 0.00360 | 0.00354 | 0.00479 |

**Fig. 2**: Normalized threshold voltage mismatch, computed with Three Points Hamer routine and maximum slope algorithm, for the same devices reported in Fig.1. 95% confidence bars are superimposed to the plot. The number of tested dice is 106.

Independently from the adopted measurement technique, a significant improvement in *K* matching properties is observed at higher temperature, while it is difficult to draw a conclusion for *Vt*, also keeping in consideration the statistical error affecting the mismatch data due to the limited sample size [4]. Fig 3 and Fig 4 report the trends of *K* and *Vt* mismatch versus *1/sqrt(WL)* for the whole set of device geometries under test: the mismatch coefficients *Avt* and *Ak* of Pelgrom model [5], i.e. $\sigma(\Delta Vt)=Avt/\sqrt{(WL)}$ and $\sigma(\Delta K/K)=Ak/\sqrt{(WL)}$, have been extrapolated, confirming the reduction of *Ak* at increasing temperature. Similar results are reported in Table I where the values of *Avt* and *Ak* at three temperatures are indicated for different device families.

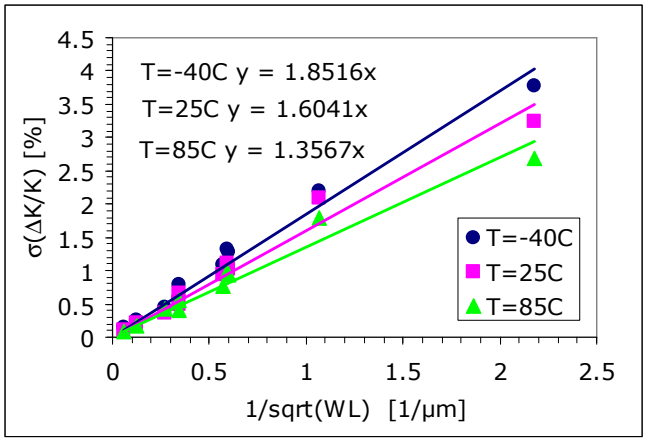



**Fig. 3**: Gain mismatch (measured through Hamer routine) versus *1/sqrt(WL)* for thick oxide N-channel mosfets at temperature T=-40C, 25C, 85C. The best fit lines and their coefficients for each temperature are superimposed to the plot. The number of tested dice is 82.

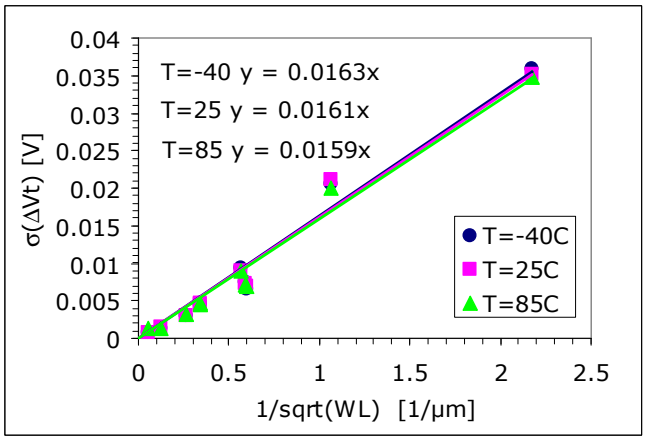



**Fig. 4**: Threshold voltage mismatch (measured through Hamer routine) versus *1/sqrt(WL)* at temperature T=-40C, 25C, 85C for the same devices of Fig. 3. The best fit lines and their coefficient for each temperature are superimposed to the plot. The number of tested dice is 82.

**Table I**: Threshold voltage and gain mismatch coefficients extrapolated from the linear fitting of experimental mismatch versus *1/sqrt(WL)* at temperature T=-40C, 25C, 85C. *Avt* is expressed in mV*μm, *Ak* is expressed in %*μm.

| DEVICE | | T=-40C | T=25C | T=85C |
|---|---|---|---|---|
| N-channel thick oxide | *Ak* | 1.85 | 1.60 | 1.36 |
| | *Avt* | 16.3 | 16.1 | 15.9 |
| P-channel thick oxide | *Ak* | 1.12 | 0.93 | 0.87 |
| | *Avt* | 7.2 | 7.1 | 7.0 |
| N-channel thin oxide | *Ak* | 2.11 | 1.63 | 1.25 |
| | *Avt* | 7.5 | 7.1 | 6.9 |
| P-channel thin oxide | *Ak* | 0.96 | 0.81 | 0.71 |
| | *Avt* | 3.5 | 3.3 | 3.2 |

To have a better insight into the phenomenon, we have also considered current mismatch. Fig. 5 reports the trend of *Id* mismatch versus overdrive for a thick oxide PMOS at two different temperatures, highlighting a significant difference at high overdrive, where the contribution of *K* mismatch is more pronounced. To summarize, *K* and *Id* mismatch at high overdrive tend to improve at increasing temperature, while *Vt* mismatch is impacted by temperature in minor quantity.

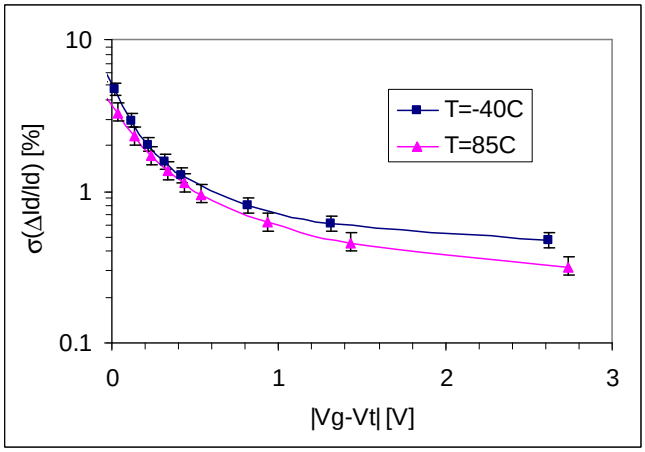



**Fig. 5**: Saturation current mismatch versus effective overdrive for a thick oxide P-channel mosfet at temperature T=-40C, 85C. 95% confidence bars are superimposed to the plot.

## CURRENT GAIN MISMATCH AND TEMPERATURE

We remind that current gain mismatch is affected by local fluctuations of device width *W* and length *L*, oxide capacitance *Cox* and mobility μ according to the basic model [5]:

$$\frac{\sigma^2(\beta)}{\beta^2}=\frac{\sigma^2(W)}{W^2}+\frac{\sigma^2(L)}{L^2}+\frac{\sigma^2(Cox)}{Cox^2}+\frac{\sigma^2(\mu)}{\mu^2} \qquad (1)$$

The local fluctuations of mobility tend to vary with temperature, due to the different impact of the scattering phenomena with temperature. To better understand this point, we recall that in the usual operating range for a real

device (between -40C and +200C, depending on the application) the low field mobility is essentially affected by the coulombian or impurities scattering (CS) and by the phonon scattering (PS) [6]. The former is dominant at low temperature, while the latter affects low-field mobility at higher temperature, as sketched in the inset of Fig. 6. The combined effect of the two scattering phenomena originates a decreasing trend of the mean value of mobility with temperature: in particular, in our devices the PMOS shows a major CS relative impact ($\mu_0(T)\sim T^{-1}$) while the NMOS is essentially dominated by PS ($\mu_0(T)\sim T^{-1.5}$) (Fig. 6). Passing from mean values to local fluctuations, we can expect that also the local fluctuations of mobility are given by the sum of two contributions related to coulombian or phonon scattering phenomena. As temperature decreases, the relative weights of mobility fluctuations due to CS or PS changes: the PS tends to freeze out, while the impurity scattering increases, causing a significant growth of mobility fluctuations [7].

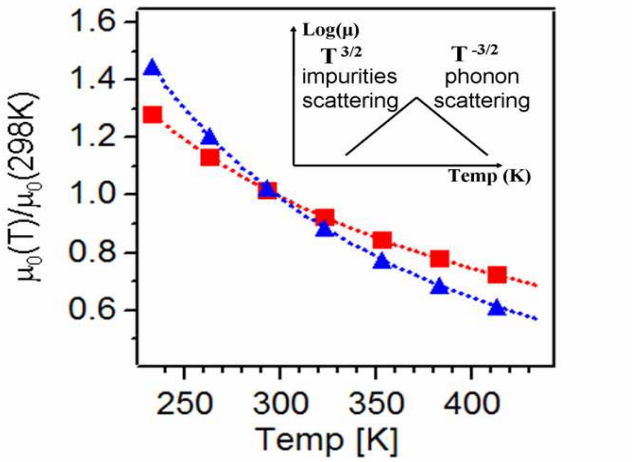



**Fig. 6:** Normalized temperature trend of the low field mobility. The N and P channel (triangles and squares) exhibit different trends. In the inset we have sketched the different temperature dependence of the phonon scattering and coulombian scattering.

Fig. 7 reports for a PMOS the trend of *K* mismatch, normalized w.r.t. device area, versus *sqrt(WL)* at three different temperatures: it suggests that the trend of gain mismatch is not linear with temperature, since the largest variation is observed in the range between -40C and 25C, rather than between 25C and 85C, even taking into account the slightly different temperature drop. Moreover it highlights that the impact of temperature is more pronounced on large area devices, where W and L fluctuations are negligible.

The trend of gain mismatch *Ak* (called $Ak_t$) versus *T-Tref* for thick oxide P-channel mosfets is shown in Fig. 8 with superimposed a simple linear fit. The linear dependence of $Ak_t$ w.r.t temperature has been introduced since the available measurements referred only to three values of T; further improvement of the model may be reached by optimizing such relationship with temperature, however also the simplified linear model appears excellent and a more complex relationship may become costly in terms of measurement time.

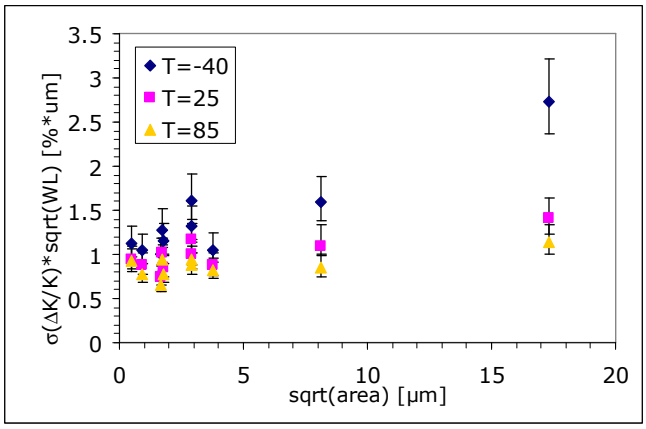



**Fig. 7**: Normalized current gain mismatch versus *sqrt(WL)* for thick oxide P-channel mosfets at three temperature values. 95% confidence bars are superimposed to the plot.

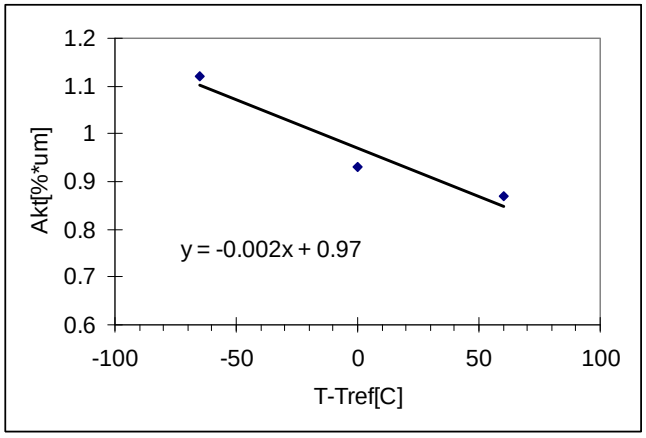



**Fig. 8**: Trend of gain mismatch coefficient *Ak* versus *T-Tref*, (Tref=25C is the reference temperature) for thick oxide P-channel mosfets. The fitting linear equation is superimposed to the plot.

## SPICE SIMULATION AND EXPERIMENTAL RESULTS

We have investigated how the temperature effect can be captured in model cards by using Monte Carlo simulations, focusing our attention on BSIM3. First of all, we have applied the usual Pelgrom square root area equation to the model parameters *u0* and *vth0*, with coefficients *Avt* and *Ak* extrapolated from the linear fitting of *Vt* and *K* mismatch versus *1/sqrt(WL)* extracted at reference temperature Tref=25C. We point out that this basic mismatch model well depicts mismatch in all the bias conditions at Tref. By performing Monte Carlo simulations we have observed that, even though the behavior of current with temperature is well described in terms of mean values, mismatch is respectively over-estimated and under-estimated at temperatures higher or lower than Tref; the differences are more pronounced at

high overdrive, thus confirming the need to better model the behavior of gain mismatch with temperature. An example is reported in Fig. 9, Fig. 10 and Fig.11 where we have plotted the trends of mean value and mismatch of *Id* versus gate bias *Vg*, both in terms of simulation and experimental results, for a PMOS at T=-40C, 25C, 85C.

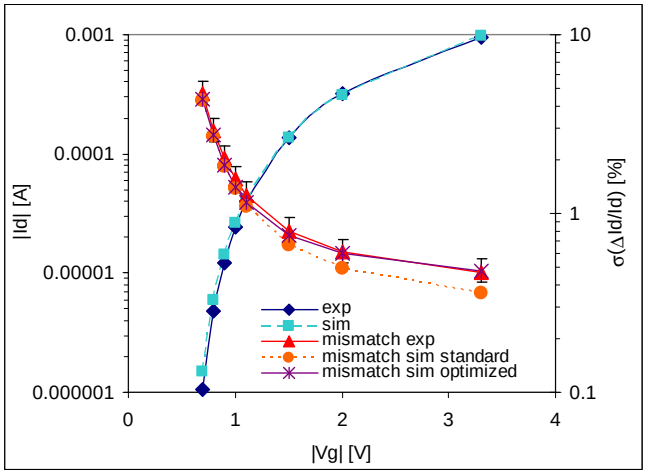



**Fig. 9**: Mean value and mismatch of saturation current versus gate bias *Vg* for a large area thick oxide P-channel mosfet form (*W*=8.4µm, *L*=1µm) at temperature T= -40C. Experimental and simulated results are compared. Simulations are realized both with standard and temperature-optimized method. 95% confidence bars are superimposed to the experimental values of mismatch.

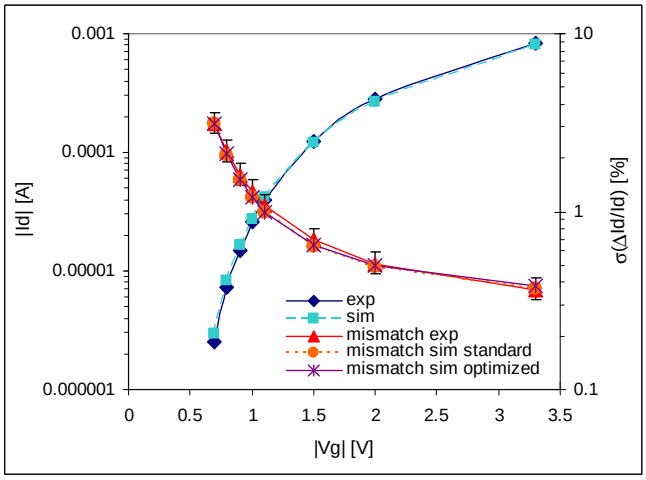



**Fig. 10**: Mean value and mismatch of saturation current versus gate bias Vg for the same device of Fig.9 at temperature T=25C. Experimental and simulated results are compared. Simulations are realized both with standard and temperature-optimized method. 95% confidence bars are superimposed to the experimental values of mismatch.

To improve modeling quality, we have modified the mismatch equation for *K* according to the following formulas:

$$Ak_t = Ak_0 + Ak_1 * (T - Tref) * \sqrt{WL} \qquad (2)$$

$$\sigma(\Delta K / K) = Ak_t / \sqrt{WL} = Ak_0 / \sqrt{WL} + Ak_1 * (T - T_{ref}) \qquad (3)$$

where $Ak_0$ and $Ak_1$ are extrapolated from a linear fit of the Pelgrom coefficients $A_k$ (extracted at different temperature) versus temperature.

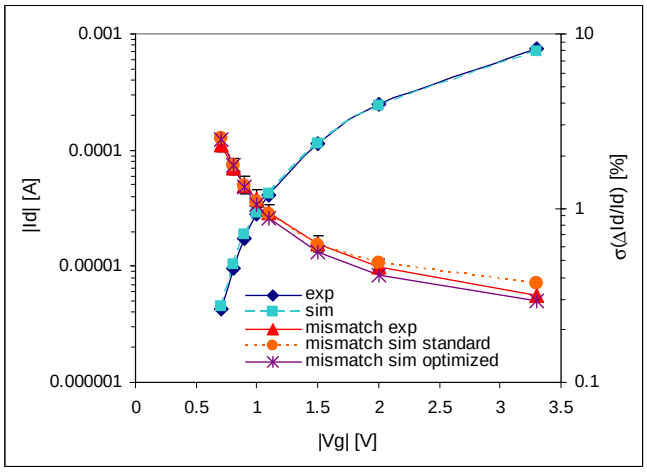



**Fig. 11**: Mean value and mismatch of saturation current versus gate bias Vg for the same device of Fig.9 at temperature T=85C. Experimental and simulated results are compared. Simulations are realized both with standard and temperature-optimized method. 95% confidence bars are superimposed to the experimental values of mismatch.

To reproduce the temperature impact dependence on device area observed in Fig. 7, $Ak_1$ has been multiplied for $\sqrt{(WL)}$. We highlight that the term $Ak_1$ may originate an unrealistic not null offset for large area devices at $T \neq T_{ref}$, in opposition with the well-known scaling law area. This is certainly a limit of our model and further analysis is necessary to better understand the role of device area and include it into a physical-based model; however our simplified model can still be considered acceptable if it is verified that, in the range of device areas and temperatures of major interest for design purpose, the term $Ak_1*(T-T_{ref})$ is significantly inferior than experimental mismatch and consequently no constant offset is originated. In the case of the devices under test, we have verified that in the temperature range between -40C and 85C the term $Ak_1*(T-T_{ref})$ is much inferior than the scaling law term $AK_0/sqrt(WL)$ up to areas of 300 µm$^2$, thus proving the validity of our model in the over cited range of device areas. Examples of Monte Carlo simulation carried out with the temperature optimized model are reported in Fig. 9, 10, 11, for a large area PMOS, and in Fig. 12, 13, 14 for a small area PMOS. Good agreement with experimental results can be observed in both the cases; furthermore, a significant improvement of the temperature optimized model is found for the large area device; while for the small area one the standard and the temperature optimized model provide similar results since, in this case, the impact of temperature on mismatch is negligible.

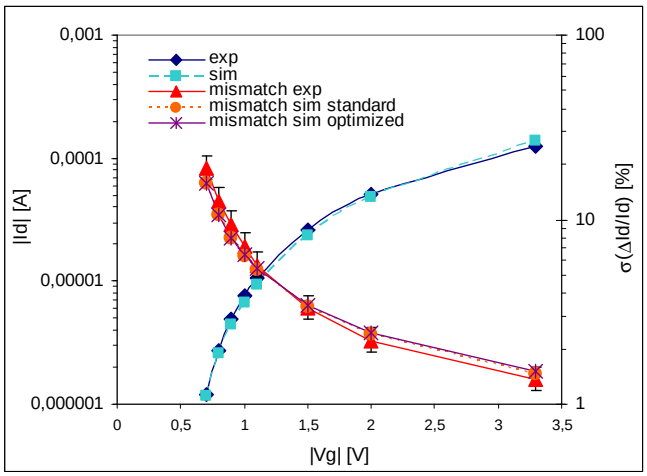



**Fig. 12**: Mean value and mismatch of saturation current versus gate bias *Vg* for a small area thick oxide P-channel mosfet from (*W*=0.48µm, *L*=0.44µm) at temperature T= -40C. Experimental and simulated results are compared. Simulations are realized both with standard and temperature-optimized method. 95% confidence bars are superimposed to the experimental values of mismatch.

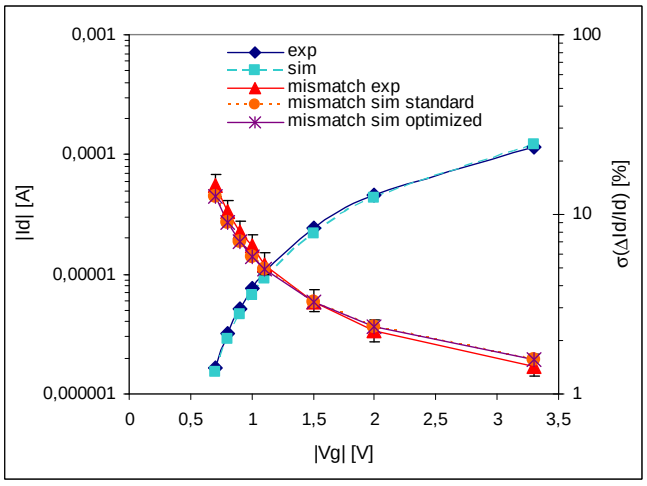



**Fig. 13**: Mean value and mismatch of saturation current versus gate bias *Vg* for the same device of Fig. 12 at T= 25C. Experimental and simulated results are compared. Simulations are realized both with standard and temperature-optimized method. 95% confidence bars are superimposed to the experimental values of mismatch.

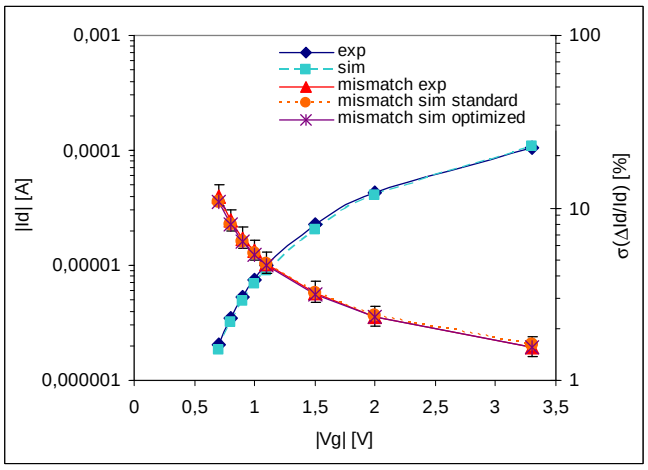



**Fig. 14**: Mean value and mismatch of saturation current versus gate bias *Vg* for the same device of Fig. 12 at T= 85C. Experimental and simulated results are compared. Simulations are realized both with standard and temperature-optimized method. 95% confidence bars are superimposed to the experimental values of mismatch.

## CONCLUSIONS

Summarizing the results collected on different technologies, we have studied the impact of temperature on MOSFET mismatch, highlighting the improvement of current gain matching properties with temperature. The possible physical explanation to this phenomenon is related to the different weight and impact on carrier mobility of the scattering phenomena with temperature (impurity scattering more evident at low temperature and phonon scattering enhanced on increasing temperature). Intuitively the difference of channel dopant in matched pairs is more visible if we emphasize the role of impurity (columbian) scattering and this can be done by reducing the phonon scattering by a temperature reduction. Finally we proposed a simple BSIM3 model implementation for Monte Carlo mismatch simulations which takes into account the temperature effect including its first order dependence on device area in a simplified way.

## ACKNOLEDGEMENTS

The authors whish to thank STM Agrate pilot line for supplying samples from different technologies for comparison purposes.

## REFERENCES


[1] P.B.Y. Tan, A.V. Kordesch, O. Sidek, “CMOS transistor mismatch model with temperature effect for HSPICE and SPECTRE”, IEEE Proceedings ICSICT 2004, vol. 24, pp. 1129-1142.

[2] X. Wu, Z. Chen, P. Madhani, “Physics and modeling of transistor matching degradation under matched external stress”, IEEE Proceedings ICMTS 2008, pp. 233-237.

[3] J.A.Croon, H.P. Tuinhout, R. Difrenza, J. Knol, A.J. Moonen, S. Decoutere, H.E. Maes, W. Sansen, “A comparison of extraction techniques for threshold voltage mismatch”, IEEE Proceedings ICMTS 2002, vol. 15, pp. 235-239.

[4] A. Pergoot, B. Graindourze, E. Jassens, J. Bastos, M. Steyaert, P. Kinget, R. Roovers, W. Sansen, “Statistics for matching”, Proceeding ICMTS 1995, pp.193-197.

[5] M. J. M. Pelgrom, A. C. J. Duinmaijer, A. P. G. Webers, “Matching properties of MOS transistors”, IEEE Journal of Solid-State Circuits, vol. 24, n. 5, 1989, pp. 1433-1440.

[6] S. M. Sze, ”Semiconductor Devices”, Wiley, N.Y., 1985.

[7] P.Y.Yu, M. Cardona, ”Fundamentals of Semiconductors”, Springer-Verlag, Berlin, 1996.

# Mismatch Measure Improvement Using Kelvin Test Structures in Transistor Pair Configuration in Sub-Hundred Nanometer MOSFET Technology

Cecilia M. Mezzomo[1,2], Mathieu Marin[1], Cedric Leyris[1] and Gérard Ghibaudo[2]

[1] STMicroelectronics
850, rue Jean Monnet 38926 Crolles, France
[2] IMEP-LAHC, Minatec
3, Parvis Louis Néel BP 257, 38016 Grenoble, France

## ABSTRACT

In this paper, a mismatch test structure in standard pair configuration using Kelvin method is introduced to better estimate the MOSFET local electrical fluctuations in sub-hundred nanometer technologies. Considering this test structure configuration, the impact of extern access connections on threshold voltage (Vt), gain factor ($\beta$) and drain current ($I_D$) mismatch extraction is investigated. To exhibit the impact of this parameter, the Vt and $\beta$ are then extracted using extrapolation method. We demonstrate that the variability of access connections does not impact Vt mismatch whereas drain current matching is underestimated without Kelvin method.

## INTRODUCTION

For 45 nm technologies and beyond, MOS transistor intrinsic variability, denominated mismatch, is becoming crucial. Mismatch consists in evaluating the electrical differences of two paired of identical transistors. Mismatch performances are important for both analog and digital circuits, such as digital-analog converter (DAC) and SRAM (respectively), since many blocks are based on the availability of pairs of electrically identical devices.

Recently, Kuroda [5] studied the characterization of transistor performance using Kelvin test structures on fully-depleted SOI CMOSFETs. Kelvin structures are currently used for arrays circuits [6]. However, mismatch test structure can be standard matched transistor pairs [4] or matrix arrays [1][7][8]. The advantage of a matrix is the large transistor population and the reduction of pads. Nevertheless, there are many disadvantages as the injection of false values due to the circuit impacts in one region of the matrix [2]. On the other hand, the pair of transistors has as advantages all the matrix arrays disadvantages (and vice-versa). These structures are simpler and its problems are well controlled. In transistor pair, the structure is designed as symmetrical as possible, at a minimum design rule, but the access to the devices, i.e. the interconnect lines or probe contact resistances, defined in this paper as "extern access connections", might be the source of not negligible voltage drops when measuring large aspect ratio (width/length) transistors. Hence, Vt matching, one of the most important parameter, can be influenced by these parasitic resistances during the extraction. However, it is not the only constraint: $\beta$ and $I_D$ are some examples of other parameters, which are probably influenced [3].

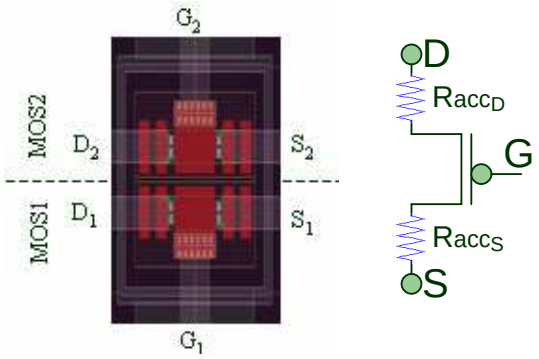
To ensure good measures, test instruments have to provide a good repeatability in order not to add extra mismatch. Thus, a routine is used to measuring the probes resistance between two touchdowns [4], controlling them during the entire test. However, this is not the only source of parasitic components. Hence, to achieve an accurate estimation of stochastic electrical parameters fluctuations, a mismatch test structure is proposed for the first time based on four-ports Kelvin method on transistor pair configuration. In addition, to correctly bias the transistors an algorithm of polarization is used. These studies are performed on 45nm standard bulk MOSFETs technology.

First, the proposed mismatch test structure is presented following by the algorithm description of how it works, i.e., the algorithm used to correctly polarize the transistor. Before presenting the results, the motivation to use these structures is clarified. Afterwards, the threshold voltage, gain factor and drain current mismatch parameters are carefully analyzed and a discussion of the measurement impacts is done.

## TEST STRUCTURE

Fig. 1.a shows conventional test structure to estimate mismatch parameters. This structure is composed of two identical MOSFET transistors (transistor pair), placed very closely, at a minimum design rule, in an identical environment. They have common bulk and separate gate

(G), drain (D), source (S) and symmetrical connections. However, one of the defaults of this structure is that it considers the extern access resistances, represented in Fig. 1.b. One resistance is the access resistance localized between the transistor channel and the drain terminal ($Racc_D$) and the other one is situated between the other side of the transistor channel and the source contact ($Racc_S$). They may be responsible of an important error introduction (as seen afterwards). To avoid this, a Kelvin structure is introduced (Fig. 2).




a) b)

Fig. 1 - Conventional mismatch test structure. a) layout and b) transistor schematic of one transistor with drain and source extern access resistances ($R_{accD}$ and $R_{accS}$, respectively).

The difference between Kelvin mismatch test structure and the conventional one is that, in the first one, drain and source have two connections called force and sense terminals. These two additional terminals allow the measurement of the effective biasing applied to the device.

To observe the impact of extern access resistances we use Kelvin mismatch test structure in three different manners:

I- as Kelvin mismatch structure: It has force and sense terminals to the drain and to the source (Fig. 2a,b). To make it available, an algorithm of polarization is applied (see next section). Hence, the effective drain and source measures (Deff and Seff) do not include the parasitic resistances (Fig. 2c).

II- as short access: Only the sense terminals are used, i. e., it takes into account the closest connections to transistor channel. In this case, the algorithm of polarization is not used (Fig. 3a).

III- as long access: Only force terminals are used, i. e., it takes into account the biggest paths. Consequently, it is more resistive than the "short" one. In this case, the algorithm of polarization is not used (Fig. 3b).

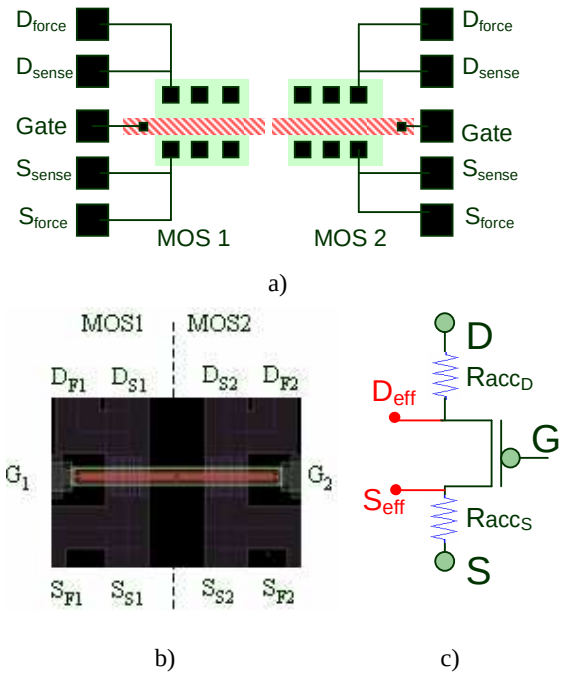



Fig. 2 - Mismatch test structure using Kelvin method. a) layout showing the terminals, b) screenshot of the real layout and c) schematic view of one transistor with forced (D and S) and sense terminals (Deff and Seff).

On Fig. 3, only one transistor of each configuration is represented.

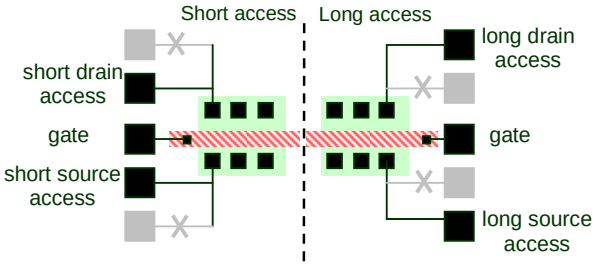



Fig. 3 – Layout of Kelvin test structure used as standard mismatch structure, i. e., it does not use all terminals. On the left side, the definition of short access and on the right, the long one.

## ALGORITHM OF POLARIZATION

The following algorithm is developed in order to adjust the polarization of the transistors on Kelvin test structure (Fig. 4).

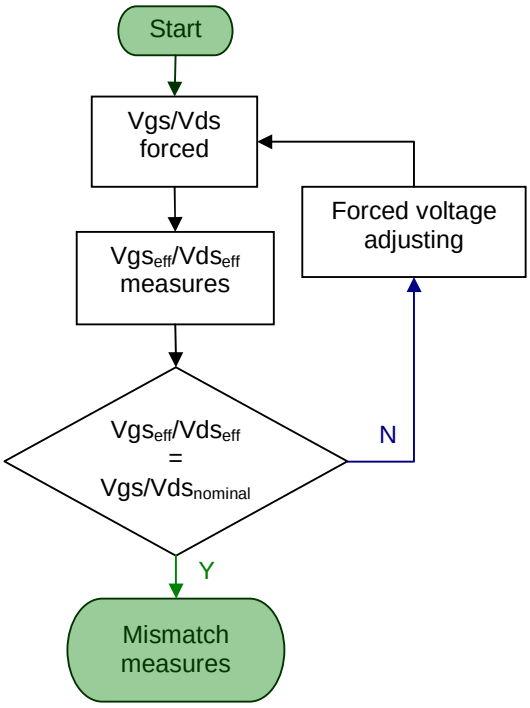



Fig. 4 - Algorithm flow to adjust forced terminals.

Nominal gate (Vgs) and drain (Vds) voltages are forced. After, the effective Vgs/Vds voltages are measured by source/drain sense terminals. Then forced terminals (first the gate and after the drain) are adjusted until electrical potential drops caused by external access resistances (probe contacts and interconnection lines) are compensated, i.e., until effective gate and drain voltage ($Vgs_{eff}$ and $Vds_{eff}$, respectively) are equal to nominal ones. Once these steps are done, Vgs and Vds are able to supply the potential drops. Therefore, mismatch parameters can be estimated.

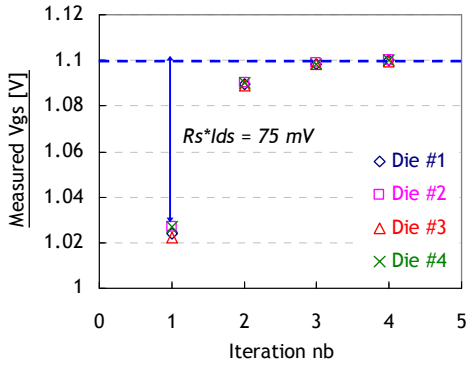



Fig. 5 – Number of iterations to adjust Vgs value in four different dices. Final Vgs voltage differs of 75mV from the nominal one.

The introduction of this algorithm in the characterization flow can increase the measurement time. However, a maximum of four iterations is needed to achieve the correct Vgs values, as shown on the example of Fig. 5. In this example, the effective gate-to-source voltage as a function of the number of iterations for different dice is presented. The final Vgs voltage differs of 75mV from the nominal one.

## EXPERIMENTAL SETUP

To ensure an acceptable statistical accuracy of matching measurements, at least 70 pairs of transistors are measured. Considering δP (δP = ΔP or ΔP/P) the difference of P parameter measured between the two paired devices, matching characterization studies result in evaluating both the mean δP and the standard deviation σ(δP) from δP Gaussian distribution.

Once the extraction is done, the first step to get rid of erroneous data is the data filtering. A recursive filter is used to eliminate all values outside the 3σ interval.

Threshold voltage (Vt) and gain factor (β) parameters are extracted by $I_D$-$V_G$ extrapolation method at maximum of trans-conductance, with $|V_G|$ varying from 0V to 1.1V by 25mV steps, and constant $|V_D|$=50mV and 1.1V.

Usual matching investigations result in the extraction of a matching parameter $A_P$, which is the proportionality coefficient between standard deviation of δP and reverse square root of drawn gate area, following the relation:

$$\sigma(\delta P) = \frac{A_P}{\sqrt{W.L}} \quad (1)$$

Matching parameter $A_P$ is obtained by calculating the mean of all the normalized mismatches:

$$A_P = \frac{1}{N}\sum_{i=1}^{N} \sigma(\delta P)_i . \sqrt{W_i.L_i} \quad (2)$$

where N is the number of transistor geometries.

In this paper, the parameters P analyzed are Vt, β and $I_D$. Measurements are performed on 45nm bulk MOSFET technology, at ambient temperature (25°C). The device under test is a NMOS with W/L ratio equals to 10µm/0.15µm.

## WHY USE KELVIN STRUCTURES

As mentioned before, Kelvin mismatch structure has more complexity than the conventional one. More run time is needed because of the algorithm used to polarize the transistors. Also, the surface used is slightly bigger due to its additional terminals. However, electrical mismatch parameters are obtained by the extrapolation method, where the extraction is based at maximum transconductance ($g_m$) [9]. The gm and the drain current $I_d$ are clearly impacted by extern access resistances as illustrated in Fig. 6. The maximum $g_m$ is reached for different Vg and its value is smaller for the conventional one. The current $I_D$ is higher using Kelvin. Consequently, these differences may provoke erroneous extractions. Vt parameter is supposed to be sensitive to them. Therefore, Vt mismatch will be investigated. It is the most important mismatch parameter, but not the only one. Hence, β factor and Id will be also studied.

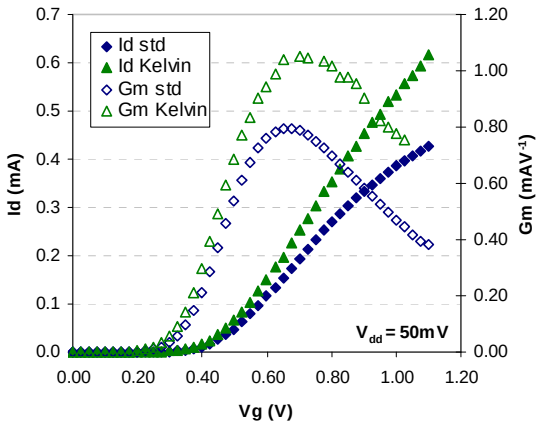



Fig. 6 – $I_D$(Vg) and gm(Vg) curves on standard and Kelvin structures. The differences represent the impact of extern access resistances.

## RESULTS AND DISCUSSION

First of all, in order to validate the functionality of the test structures, tests of repeatability have been performed in a different bench to verify if there are no extern influences. These tests were verified on proposed structures (Kelvin, long and short). The results obtained are in accordance with the initial ones.

On Fig. 7 is presented the cumulative distribution of Vt. It points out that Vt values are not the same for long, short and Kelvin mismatch structures. This demonstrates that Vt is impacted with extern access resistances. The long structure has the lowest Vt because it has the highest access resistance, thus the highest potential drop. Consequently, the Vt obtained using standard structure is not a true value. The correct one is given with Kelvin structures, which value is higher than the other ones.

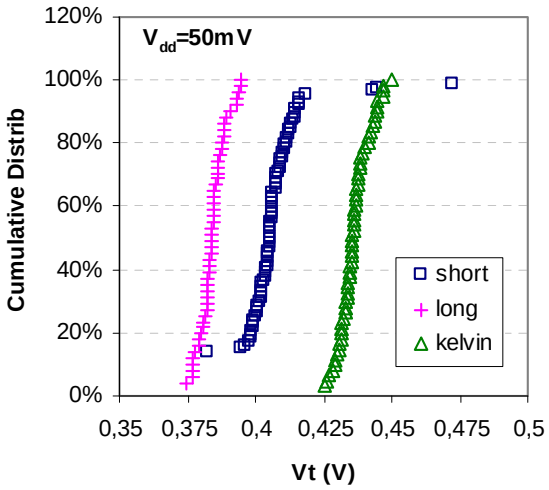



Fig. 7 - Cumulative distribution of Vt for short and long access and Kelvin structures.

Fig. 8 shows Vt values of the transistor pair (MOS1 and MOS2) for the three studied test structures. It is possible to remark that Vt values of MOS1 are not exactly the same of MOS2, indicating that Vt mismatch exists. The mismatch depends on the Vt difference measured between the transistors pair (ΔVt) and not really of the intrinsic Vt value. Hence, although there is Vt shift between the different test structures, it is not sure that it impacts the mismatch. Therefore, Vt mismatch is evaluated in following.

Thus, the impact of extern access connections on Vt mismatch will be discussed. On Fig. 9 is shown the cumulative distribution of Vt variations.

Although there is a Vt shift between the test structures, the mismatch of Vt is not impacted. It is possible to notice the three ΔVt curves corresponding to short, long and Kelvin structures are superposed. Also, the cumulative distribution is centered in zero. It is demonstrated that the variability of access connections does not impact significantly the Vt mismatch as transistors are symmetrical. This is true if the pair of transistors is identical including the connections at an identical environment. Consequently, conventional test structures, which have less complexity and are faster than Kelvin, can be used to measure Vt mismatch.

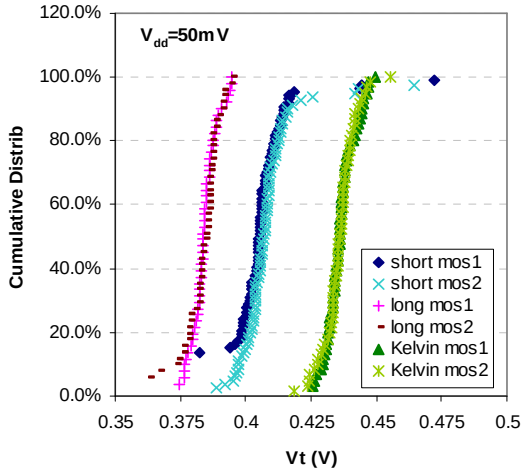



Fig. 8 - Cumulative distribution of Vt for transistor 1 (MOS1) and transistor 2 (MOS2) to short and long access and Kelvin structures.

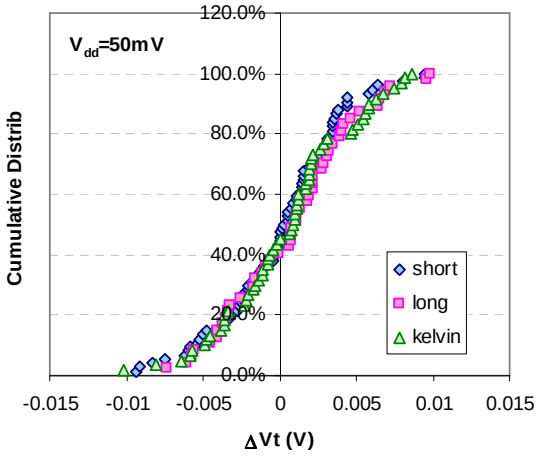



Fig. 9 - Cumulative distribution of ΔVth for short and long access and Kelvin structures.

However, the same conclusion as Vt mismatch could not be made for β factor. In this case, the mismatch of β is more delicate as shown in Fig. 10. Different results are obtained when using the short access, the long access or the Kelvin method. The short one is not centered in zero and Kelvin has not a good performance also. Therefore, these problems may be crucial especially in analog circuits. Therefore, more studies are necessary to enhance reliability in the extraction of this parameter.

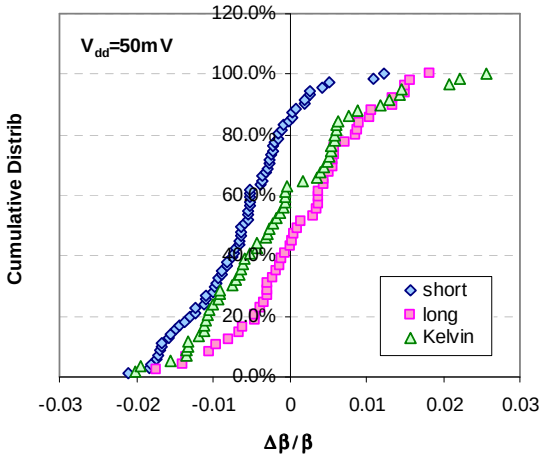



Fig. 10 - Cumulative distribution of Δβ/β for short and long access and Kelvin structures.

To finish the discussion, the last parameter which will be analyzed is the drain current $I_D$. Fig. 11 shows the relative mismatch of $I_D$ from weak to strong inversion at Vds = 1.1 V. As $I_D$ increases, we observe that the more resistive is the access connection to the device, the more we underestimate the drain current matching when not using the Kelvin method.

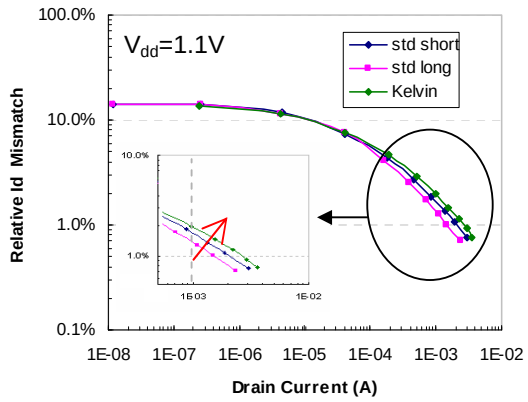



Fig. 11 - Relative $I_D$ mismatch using short and long access and Kelvin structures.

These three mismatch parameters can be incorporated into models to help designers to observe dispersions without strong mistakes. They have many applications, as in digital memories and in analog circuits with current mirrors, current sources and differential pairs. Then, β and $I_D$ mismatch are also significant technological

indicators. As in a DAC, the output in the analog circuit is provided by currents, using current mirror circuits. For example, if there is a digital signal of 8 bits, the analog circuit will need 256 current mirrors to realize de conversion. Hence, the reproducibility of the current mismatch parameters is very important.

## CONCLUSION

We developed a Kelvin test structure for MOSFET mismatch measurement in pair configuration. We showed that extern access connections do not affect the mismatch of threshold voltage Vt, even if the extracted Vt is not a true value with standard structures. On the other hand, the mismatch of β has been shown to depend significantly on access layout, introducing uncertainty in its determination. Moreover drain current matching is underestimated without Kelvin method.

## REFERENCES


[1] H. P. Tuinhout, "Design of matching test structures", ICMTS Proceedings, p. 21-7, 1994.

[2] "Characterization and modeling of local electrical fluctuations in sub-65 nm CMOS transistors", Ph. D., A. Cathignol,2008.

[3] P. Kinget, "Device mismatch and tradeoffs in the design of analog circuits", IEEE Journal of Solid-State Circuits, v. 40, issue 6, p. 1212 – 1224, 2005.

[4] A. Cathignol, S. Bordez, K. Rochereau, G. Ghibaudo, "From matching test structures to matching data utilization: not an ordinary task", ICMTS Proceedings, p. 230-3, 2007.

[5] R. Kuroda et al, "Characterization of MOSFETs intrinsic performance using in-wafer advanced kelvin-contact device structure for high performance CMOS LSIs", ICMTS Proceedings, p. 155-159, 2008.

[6] R. Lefferts, C. Jakubiec, "An integrated test chip for the compete characterization and monitoring of a 0.25 CMOS technology that fits into five scribe line structures 150µm by 5,000µm", ICMTS proceedings, p. 59-63, 2003.

[7] T. Chagawa, K. Terada, J. Xiang, K. Tsuji, T. Tsunomura, A. Nishida, "Measurement of the MOSFET drain current variation under high gate voltage", ICMTS Proceedings, p. 86-89, 2008.

[8] Y. Shimizu, M. Nakamura, T. Matsuoka, K. Taniguchi, "Test structure for precise statistical characteristics measurement of MOSFETs", ICMTS Proceedings, p. 49-54, 2002.

[9] Croon J. A. et al., "A comparison of extraction techniques for threshold voltage mismatch", ICMTS Proceedings, p. 235, 2002.

# A Test Structure for Spectrum Analysis of Hot-Carrier-Induced Photoemission from Scaled MOSFETs under DC and AC Operation

T. Matsuda[1], T. Maezawa[1], H. Iwata[1], T. Ohzone[2]

[1] Department of Information Systems Engineering, Toyama Prefectural University
Kurokawa Imizu-Shi, Toyama, Japan
[2] Dawn Enterprise, Nagoya, Japan

## ABSTRACT

A test structure with a wide channel width for analysis of hot-carrier-induced photoemission is presented and spectrum changes for 90 nm MOSFETs under DC (direct current) and AC (alternating current) operation are discussed. Comparing with DC operation, photon counts for higher photon energy increase under AC operation, and spectrum curves change with rise and fall time of gate pulse. The overshoots of drain voltage at the transition timing generate hot carriers with higher energy due to large electric field near drain region, which raise a possibility of a reliability issue related to hot carrier effects in LSIs.

## INTRODUCTION

As the channel length of CMOSFETs has been scaled down to sub-100 nm level, hot-carrier induced degradation is still one of the important problems affecting reliability. Since the hot carriers emit photons, analyses using a photoemission microscope become useful to study the high electric fields in MOSFETs [1-3]. Several models of hot-carrier-induced photoemission are proposed, such as recombination of hot electrons with holes generated by the impact ionization, bremsstrahlung radiation of hot carriers. Since digital LSIs are operated with AC (alternating current) mode rather than DC (direct current) mode, detailed hot carrier analysis under AC operation is important for scaled CMOSFETs. In this paper, spectrum measurement results of hot-carrier-induced photoemission for 90 nm MOSFETs under DC and AC operation are discussed using a specially designed test structure, which has a wide channel width for sufficient photoemission intensity.

## EXPERIMENTS

### A. *Test Structure*

Test devices of n-MOSFETs were fabricated by a standard 90 nm CMOS process with 6-level metal layers. Fig. 1 shows a photomicrograph of a test device and its schematic of array structure. The array has 7 x 22 of unit MOSFETs with 3 μm gate width connected in parallel, and thus all test structures have a total channel width W of 462 μm. Four kinds of test structures were designed with different gate length L of the unit MOSFET of 0.10, 0.11, 0.14 and 0.20 μm, where L is the drawn value in CAD and 0.10 μm corresponds to 90 nm on wafer. Fig. 2 shows a schematic cross-section of the test structure. The distance d between the polysilicon gate and the source/drain metal electrodes was designed 1.5 μm to avoid distortion of the photoemission intensity due to the reflectance/interference effects of metal electrodes.

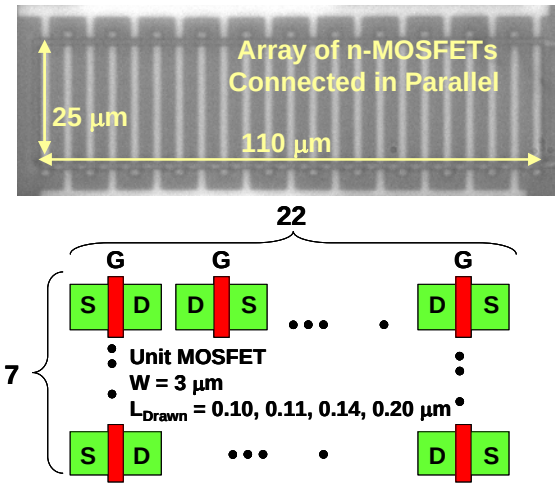



Fig. 1. Photomicrograph of a test device and its schematic of array structure.

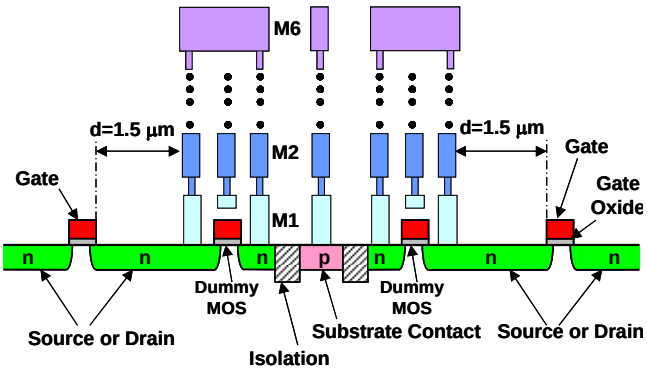



Fig. 2. Schematic cross-section of the test structure. The distance d between the polysilicon gate and the source/drain metal electrodes was designed 1.5 μm to avoid the reflectance/interference effects of metal electrodes.

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

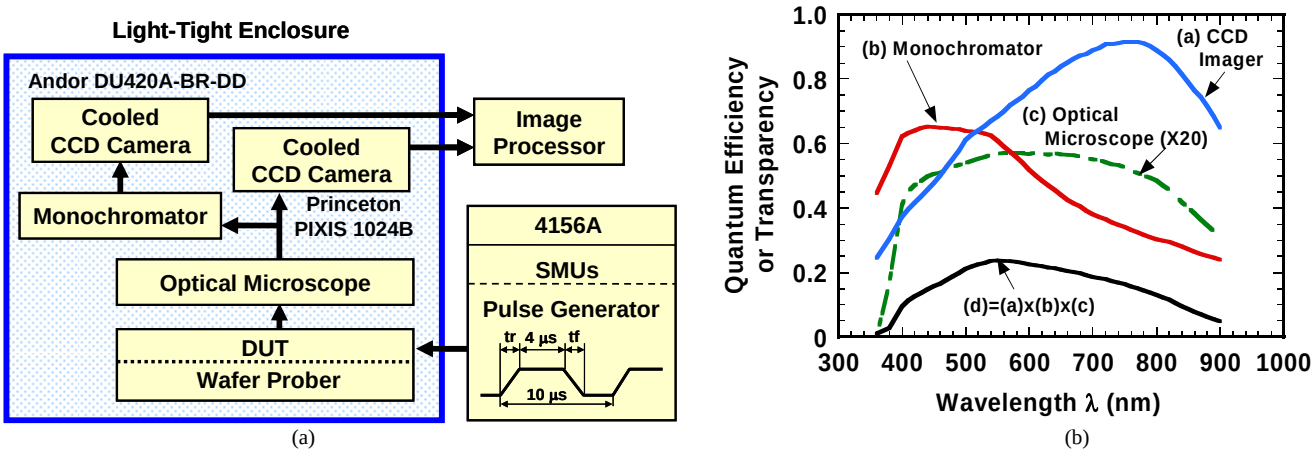



Fig. 3. (a) Schematic of the photoemission measurement system, and (b) spectral response curves of the system.

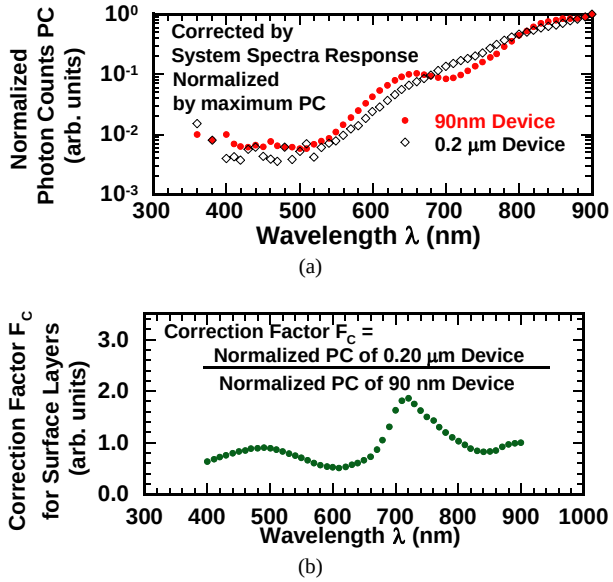



Fig. 4. (a) Normalized spectral curves of 90 nm and the previous 0.2 µm device under DC operation, (b) the correction factor $F_C$ calculated from (a).

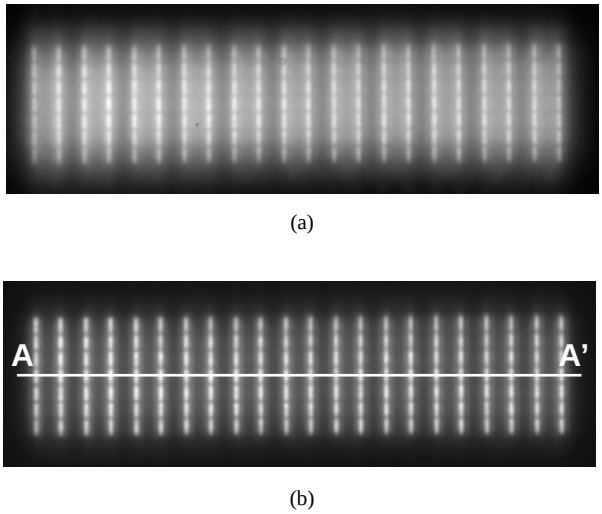



Fig. 5. Measured photoemission images under (a) DC and (b) AC operation of L = 0.10 µm devices.

### *B. Measurement System*

Hot carrier induced photoemission under DC and AC operation were measured. Gate voltage $V_G$ were kept constant at 1.0 V with $V_D$ = 1.4 ~ 2.0 V for DC operation. For AC operation, two kinds of pulse signals of $t_r$ (rise time) and $t_f$ (fall time) were applied to gate electrode keeping $V_D$ = 1.2 V. Voltage swing, period and width of $V_G$ pulse are 0 to 1.0 V, 10 µs and 4 µs, respectively. Fig. 3 shows (a) a schematic of the photoemission measurement system, and (b) spectral response curves of the system. The spectrum range from 350 to 900 nm can be measured at once with the CCD cooled at –70 °C by a Peltier device. The spectrum curves were corrected for the spectral response curve of the measurement system and surface layers. Since the surface of the test devices is covered by multi-layered thin films such as low-k and etch-stop layers, interference of reflected light among the layers affect spectra. Since precise calculation of spectral for 6 metal process response is difficult, correction factor $F_C$ for surface layers were estimated from the previous data of our 0.2 µm CMOS devices. Spectral curves of 90 nm (L=0.10 µm) and the previous 0.2 µm device under DC operation, which are normalized by the maximum values, are shown in Fig. 4(a). Since both spectra have the similar slope except a hump around 630 nm caused by the interference of multiple reflection, the correction factor $F_C$ was calculated as shown in Fig 4(b) and was applied to all measured data.

## EXPERIMENTAL RESULTS AND DISCUSSION

Fig. 5 shows measured photoemission images under (a) DC and (b) AC operation of L = 0.10 µm devices. Stripes of light aligned in 7 x 22 matrix correspond to the unit MOSFETs of W=3 µm. The DC operation gives more light diffusion from gate region than AC operation, and it suggests spectrum change of photoemission. Fig. 6 shows a photon counts PC profile sliced along the line A-A' in Fig. 4(b).

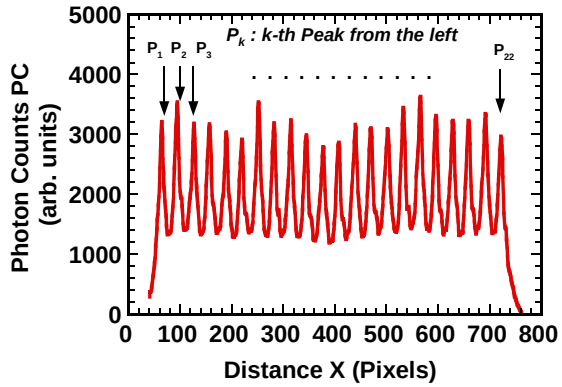



Fig. 6. A photon counts PC profile sliced along the line A-A' in Fig. 5(b).

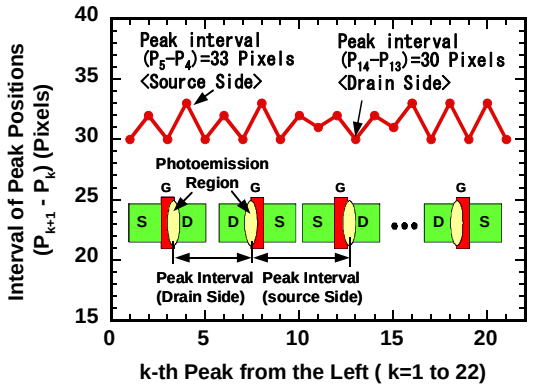



Fig. 7. Intervals between adjacent peaks ($P_{k+1}$ - $P_k$ ; k=1 to 22), where $P_k$ denotes the position of k-th peak from the left in pixels in Fig. 6.

Fig. 7 shows the intervals between adjacent peaks ($P_{k+1}$ - $P_k$ ; k=1 to 22), where $P_k$ denotes the position of k-th peak from the left in pixels in Fig. 6. The interval between adjacent peaks changes from 30 to 33 pixels by turns. Since drain and source are placed alternately, the interval variation indicates that photoemission regions shift to drain side from the gate center.

Fig. 8 shows a logarithmic plot of photon counts versus photon energy of DC operation at various $V_D$. Linear relations below 2.5 eV are similar to the published data [1], and suggest that photon energy has a Boltzman distribution; $\exp(-h\nu/kT_e)$, where $T_e$ is the electron temperature. Although the photon counts for the energy range over 2.5 eV tends to be saturated, the photon counts were small and it needs more precise measurements.

Fig. 9(a) shows gate length L dependences of hot-carrier-induced photoemission spectra under AC operation with rise and fall time $t_r$ = $t_f$ = 1 μs. Comparing with DC operation results in Fig. 8, photon counts for higher photon energy increase. Although photon counts decrease as L becomes longer, the spectrum curves give similar slopes. Fig. 9(b) shows the effect of gate length on spectrum curves of photoemission under AC operation with $t_r$ = $t_f$ = 0.1 μs. Further increase of photon counts for higher photon

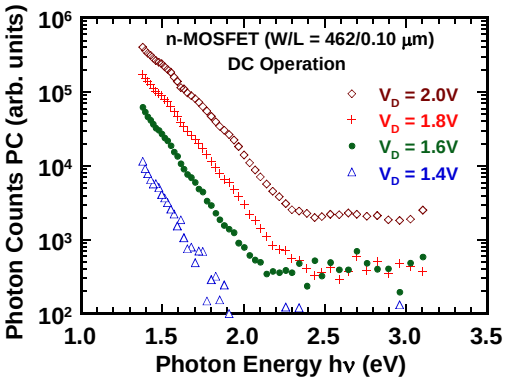
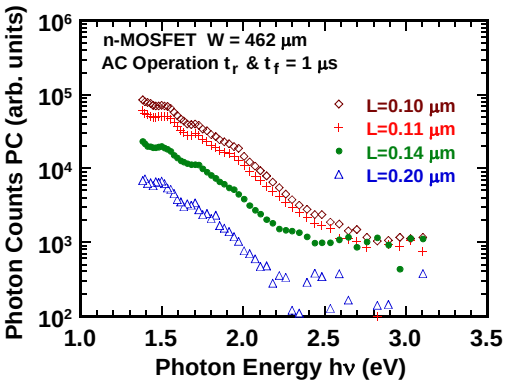
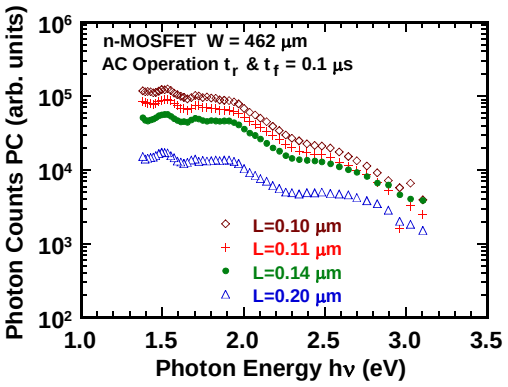



Fig. 8. A logarithmic plot of photon counts versus photon energy of L = 0.10 μm device under DC operation at $V_D$ = 1.4~2.0 V, $V_G$ = 1.0 V and $V_S$ = $V_B$ = 0.0 V.




(a)




(b)

Fig. 9. Gate length L dependences of hot-carrier-induced photoemission spectra under AC operation with rise and fall time (a) $t_r$ = $t_f$ = 1 μs and (b) $t_r$ = $t_f$ = 0.1 μs.

energy is observed. The growth of shorter wavelength components corresponds to the change of images in Fig. 5.

The spectra of DC operation and AC operation with $t_r$ ($t_f$) = 1 and 0.1 μs for L=0.10 μm device are compared in Fig. 10. Photon counts for photon energy above 2 eV under AC operation becomes larger than DC operation by 1~2 order of magnitude. Assuming a Boltzman distribution, the calculated electron

temperature $T_e$ of 3240 or 5260 K for AC operation is much higher than that of 1790 K for DC operation.

Fig. 11 shows voltage waveforms measured by an oscilloscope under AC operation. Relatively large capacitances and inductances in the test devices causes overshoots of drain voltages, which becomes larger as $t_r$ and $t_f$ of gate voltage pulse get steeper. The peak drain voltage for gate pulse of $t_r = t_f$ =0.1 μs reaches about 2.6 V. Consequently, large electric field near drain region generates hot carriers with higher energy at the transition timing [4]. It implies that overshoots of digital signals in LSIs affects reliability with hot carrier effects.

## Conclusion

A test structure for 90 nm MOSFETs, which has a wide channel width of 462 μm, has been designed, and successfully measured photoemission spectra under DC and AC bias conditions. The increase of photon counts for higher photon energy due to the overshoots of drain voltage under AC operation has been observed, and it may affect reliability of LSIs.

## Acknowledgment

The VLSI chip in this study has been fabricated through the chip fabrication program of VLSI Design and Education Center(VDEC), the University of Tokyo, with the collaboration by STARC, Fujitsu Limited, Matsushita Electric Industrial Company Limited., NEC Electronics Corporation, Renesas Technology Corporation, and Toshiba Corporation.

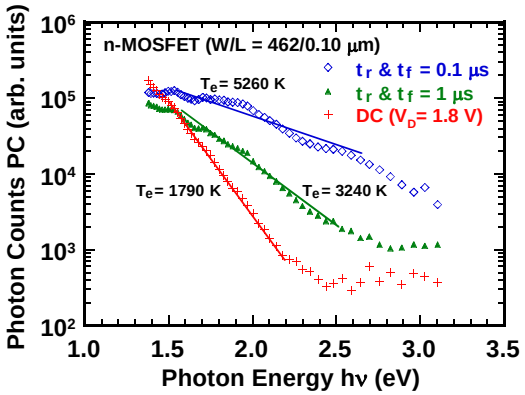



Fig. 10. Spectra of DC operation and AC operation with $t_r$ ($t_f$) = 1 and 0.1 μs for L=0.10 μm device.

## References

[1] T. Matsuda, T. Ohzone, E. Kameda S. Odanaka, K. Yamashita, N. Koike and K. Tatsuuma, "A test structure for spectrum analysis of hot-carrier-induced photoemission from MOSFETs," IEEE Transactions on Semiconductor Manufacturing, vol. 16, no.2, pp. 233-238, 2003.

[2] T. Matsuda, T. Tanaka, H. Iwata, T. Ohzone, K. Yamashita, N. Koike and K. Tatsuuma, "A test structure for spatial analysis of hot-carrier-induced photoemission in n-MOSFETs," Proc. IEEE Int. Conf. on Microelectronic Test Structures, pp.47-51 2005.

[3] T. Matsuda, H. Takeuchi, A. Muramatsu, H. Iwata, T. Ohzone, K. Yamashita, N. Koike and K. Tatsuuma, "A test structure for two-dimensional analysis of MOSFETs by hot-carrier-induced photoemission," IEICE Transactions on Electronics, vol.E88-C, no.5, pp.811-816, 2005.

[4] B. Marchand, B. Cretu, G. Ghibaudo, F. Balestra, D. Blachier, C. Leroux, S. Deleonibus, G. Guégan, G. Reimbold, S. Kubicek, K. DeMeyer, “Secondary impact ionization and device aging in deep submicron MOS devices with various transistor architectures,” Solid-State Electronics, vol. 46, pp. 337-342, 2002.

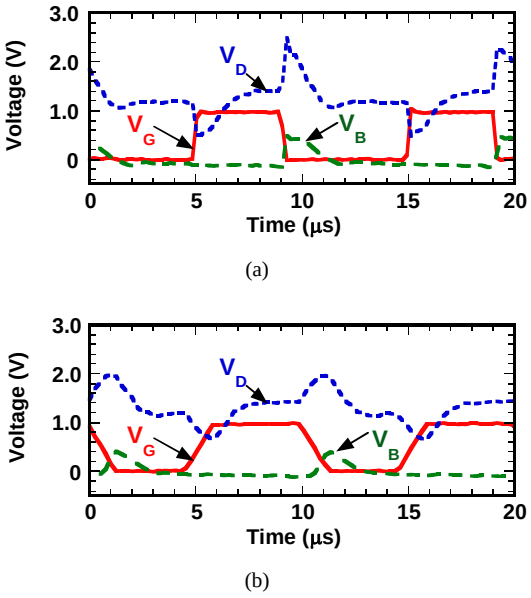



Fig. 11. Voltage waveforms measured by an oscilloscope under AC operation for (a) $t_r = t_f$ = 0.1 μs and (b) $t_r = t_f$ = 1 μs.

# Application of a Micromechanical Test Structure to the Measurement of Stress in an Electroplated Permalloy Film

S. Smith*, N.L. Brockie*, J.G. Terry*, N. Wang*, A.B. Horsfall† and A.J. Walton*

* Institute for Integrated Micro and Nano Systems, School of Engineering,
Scottish Microelectronics Centre, The University of Edinburgh, EH9 3JF, UK
Email: Stewart.Smith@ed.ac.uk

† School of Electrical, Electronic and Computer Engineering, Merz Court,
Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

## ABSTRACT

Suspended microrotating test structures designed to measure the stress in thin, surface micromachined films have been applied to the production of thick layers of electroplated permalloy (NiFe alloy). This process has particular significance to the production of magnetic MEMS components and devices. It is extremely important to characterise the stress in such materials, especially where these films are to be used on wafers with underlying integrated circuitry as it is well known that the matching of transistors can be affected by mechanical strains induced by interconnect features running above them. A new test chip has been designed and fabricated in order to determine the optimum dimensions for permalloy stress sensor structures.

## INTRODUCTION

Characterisation and control of the intrinsic stress in deposited films is important for many microfabrication processes. The micro-mechanical stress test structures used in the present work have been demonstrated with a number of different thin film processes and materials. Permalloy is a generic term for nickel iron magnetic alloy (80% Ni, 20% Fe), which is becoming more and more widely used in the production of magnetic MEMS (MicroElectroMechanical Systems) devices and integrated microscale inductors, as it can greatly improve their operation. The permalloy is deposited by an electroplating process and, as with most Nickel based materials deposited in this way, significant stresses can be induced, especially in the thick layers required for MEMS devices. The stresses in blanket deposited layers can be assessed through measurement of wafer bow, averaged across the substrate, while the stress sensor structure used here has the potential to allow local variations in stress to be mapped across a wafer with high resolution.

## TEST STRUCTURES

The microrotating stress sensor [1], [2] used in this study consists of three elements fabricated in the thin film layer being measured. These are two stress or expansion arms which anchor a pointer arm as shown in Fig. 1. In interconnect processes such as those investigated in [3], the dielectric under the metal acts as a sacrificial layer. When this is etched away it releases the suspended structure and also releases any residual stress in the expansion arms. They expand or contract causing the pointer arm to rotate. In the structure in Fig. 1, a clockwise rotation indicates compressive stress while an anti-clockwise rotation indicates tensile stress.

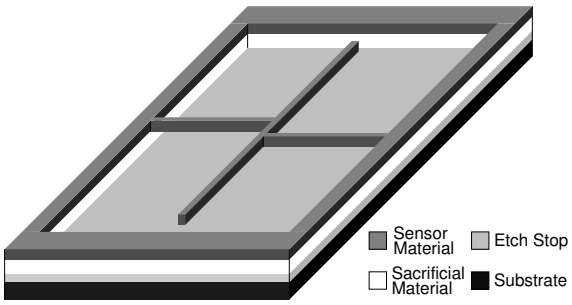



Fig. 1. Schematic diagram of a stress test structure

Similar sensors have also been used to investigate stresses induced by electromigration in interconnect. The test structures have been altered so that the anchors for the expansion arms are also pads to allow electrical connections to be made [4]. Fig. 2 shows one such structure fabricated in aluminium. Electromigration induced stress can be studied by passing current along a long chain of these structures at elevated temperatures and measuring changes in rotation from one end to the other.

## TEST CHIP DESIGN

In order to obtain the greatest sensitivity from the test structures it is important to determine the geometry which will give the maximum rotation for a given stress. Fig. 3 shows the layout of a stress sensor structure and indicates the important parameters. These are the width (W) and length (L) of the expansion arms and the separation ($\Delta$Y) between them, where they attach to the rotating pointer.

A new stress sensor test chip has been designed which will allow the optimum geometry to be determined for much thicker layers that have been previously reported with structures similar to that illustrated in Fig. 3. Three different

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

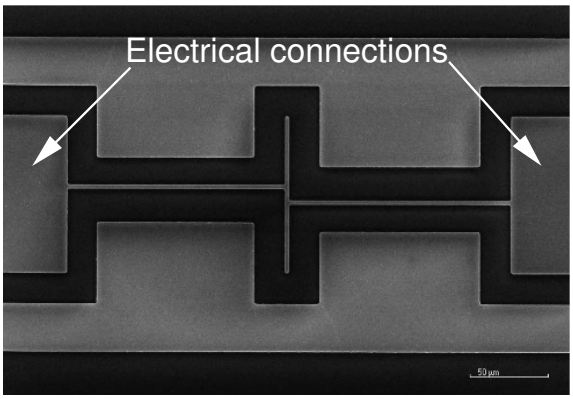



Fig. 2. SEM image of a stress sensor structure with electrical connections

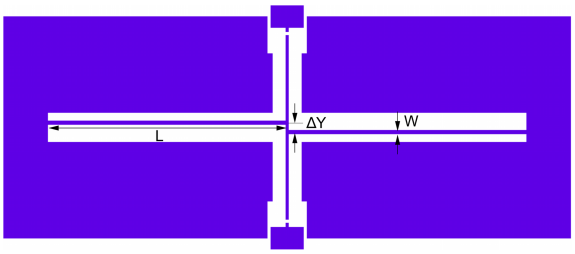



Fig. 3. Layout of stress sensor test structure

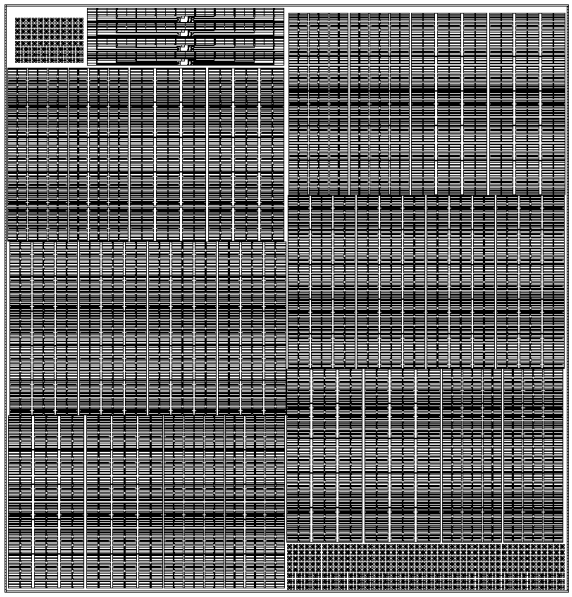


Fig. 4. Layout of stress sensor test chip

expansion arms widths (5µm, 8µm and 10µm) and lengths (650µm, 750µm and 850µm) have been used. The expansion arm separation ratio (ΔY/W) was varied from 0.5 to 5 and the full chip layout is presented in Fig. 4. The complete test chip measures around $50\times50\text{mm}^2$ and has been tiled 4 times on a 5" mask plate for contact printing.

In addition to the stress sensor structures described above, the test chip also includes Greek cross sheet resistance test structures [5] and cross-bridge electrical linewidth structures [6]. These will be used to measure the resistivity of the permalloy and to look at variation of the feature sizes across the wafer.

## PROCESSING

The permalloy test structures were fabricated on 200mm silicon wafers pre-coated with a Ti-Cu-Ti seed layer stack. The wafers were spin coated with a 28µm thick layer of Futurex NR2 photoresist before soft baking and a 10mm edge bead removal using PGMEA (Propylene Glycol Monomethyl Ether Acetate) on a Brewer Science CEE 6000 coater system. The clearing of the resist from the rim of the wafer was essential to ensure electrical contact to the copper seed layer for electrodeposition.

NR2 is a negative photoresist that was printed with the test chip pattern using a Karl Suss MA8 contact aligner. The wafer was then post exposure baked to fix the pattern and developed to remove the un-exposed resist. Once the areas for deposition had been cleared of titanium with a short etch in 1% hydrofluoric acid, the wafer was dipped in a 5% sulphuric acid bath to remove surface copper oxide before being rinsed with de-ionised water and placed in a Semitool Raider M plating tool containing a nickel/iron chloride electrolyte. The permalloy metal ($Ni_{80}Fe_{20}$) was deposited using a pulse plating process to give a final thickness of approximately 5µm.

After plating the wafer was rinsed and dried before the remaining resist was stripped using a photoresist remover solution and ultrasonic agitation. The next step was to strip the copper and titanium seed layers to allow access to the silicon substrate for the release of the test structures. The Cu seed was removed by wet etching in a 10:1, acetic:nitric acid solution until the seed could be visually observed to have been removed. An unfortunate side effect of the copper etch was that it also attacked the permalloy leading to the staining seen in some of the images presented in this paper. The final layer of Ti seed was removed by a short dip in 1% HF to expose the silicon substrate.

As the structures were fabricated directly onto the silicon rather than on a dielectric layer they were released using a $XeF_2$ vapour etch process which isotropically etches silicon. This process was performed using a Memsstar SVR XF etch release tool. Typically, this process involves a sacrificial layer of polycrystalline silicon but in this case we are etching into the bulk silicon substrate. The etch rate is difficult to define, as it depends strongly on the exposed silicon surface area, but a camera with a macro lens can be used to watch the structures and determine when they are released. The etch release process is illustrated in Fig. 5.

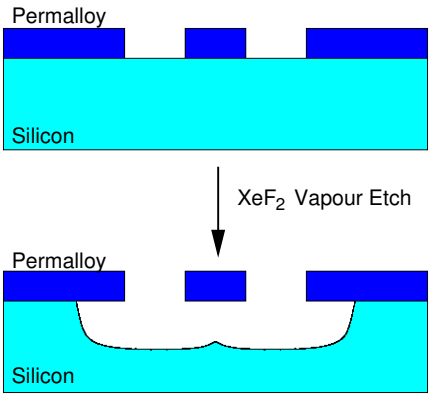
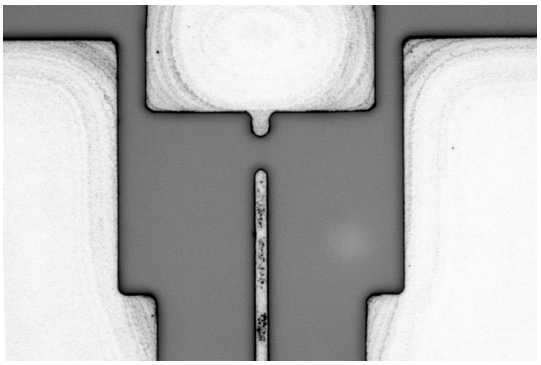
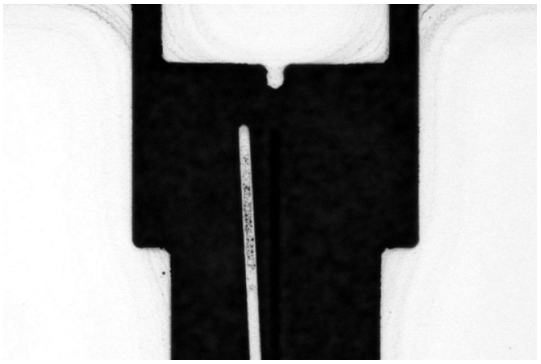



Fig. 5. Cross section through suspended structure showing $XeF_2$ vapour etch release process.



(a) Before release



(b) After release

Fig. 6. Microscope images of microrotating stress sensors

## RESULTS

Rotation of the pointer arms in the stress sensor was assessed by taking images of the released structure with a digital camera attached to an upright optical microscope and extracting the angle from the captured image. Fig. 6 shows microscope images of stress sensor structures before and after releasing. The anti-clockwise rotation after release shows that there is tensile stress in the permalloy film as expected.

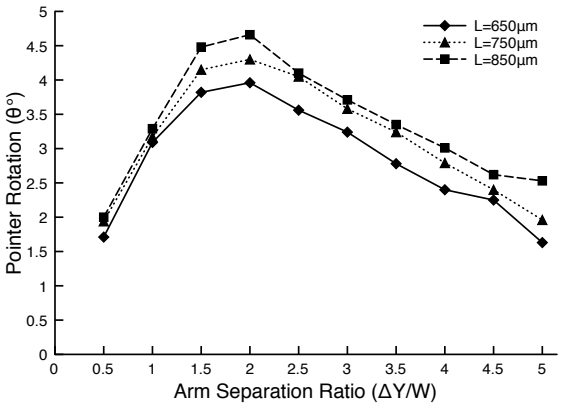
In order to properly assess the rotation of the pointer arms after release the images were analysed using the GNU Image Manipulation Program (GIMP). Any rotation initially present in the image was removed and a edge detection filter used to extract the outline of the structure. Unfortunately the test structures with expansion arm widths of 5µm did not survive to the end of processing and so the results presented here are from structures with W = 8 and 10µm. Measurements were made of two complete sets of these structures, one close to the centre of the wafer and one nearer to the edge of the printed area. The results can be seen in Fig. 7 and 8

In general, the results show that the rotation initially increases with $\Delta Y/W$, reaches a maximum and then decreases. This behaviour has a number of different possible causes including: plastic deformation of the expansion arms in the hinge region at low separations, reduction of torsional force as the separation increases, stress related stiffness in the expansion arms counteracting the rotational force and issues with the pattern fidelity where the expansion arms attach. Previously published results for aluminium suggest that $\Delta Y$ should be at least twice the arm width and this is confirmed for these structures where the maximum rotation is observed for $\Delta Y/W=2$ [3].

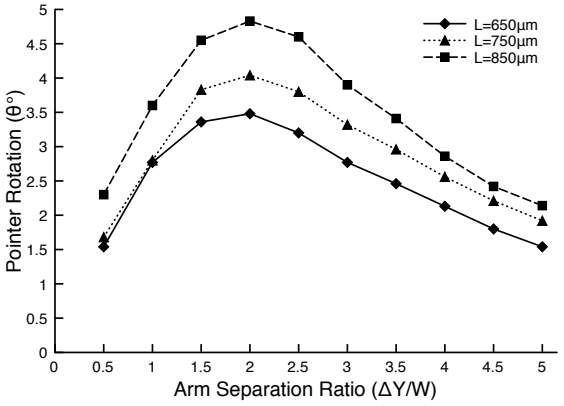
Looking at the results in more detail, those from the 8µm structures show similar curves for each of the arm lengths and an increase in rotation with arm length (L). In addition, the rotations observed on the structures nearer the edge of the wafer are smaller than those from the middle. However, it should be noted that the permalloy was thicker in the middle of the wafer and this may have contributed to this observation. There are also differences in the amount of over-etching of the test structures caused by the copper seed etch. This means that the structures are significantly narrower than designed in some parts of the wafer.

These effects may also explain some of the results obtained from the 10µm structures which do not show the same consistent trends. The results taken from nearer the middle of the wafer (Fig. 8(a)) show smaller rotations for the 850µm expansion arms than expected, even dropping below that observed for the 650µm structure where $\Delta Y/W=2$. The results in Fig. 8(b) are similar for the structures with L=650µm, which are initially greater than those from the other two sets of structures but drop below them, to values that might have been expected, as $\Delta Y/W$ increases.

Figure 9 compares the results obtained from structures with expansion arm lengths of 750µm. It suggests that, in general, the rotations are larger for the structures with narrower expansion arms and for those closer to the middle of the wafer. However, this may not be due to a difference in the stress across the wafer as there is also significant variation of the dimensions of the test structures with position. Initial measurements of a small number of the electrical linewidth

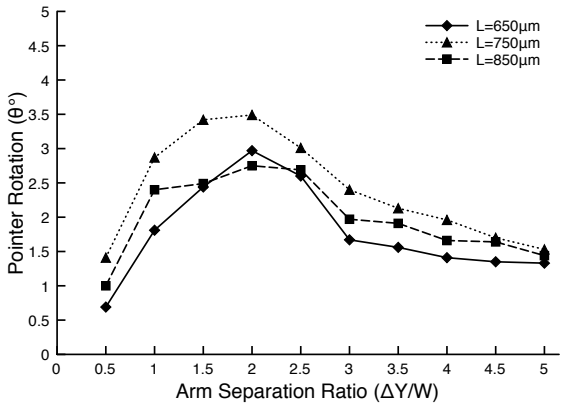



(a) Middle

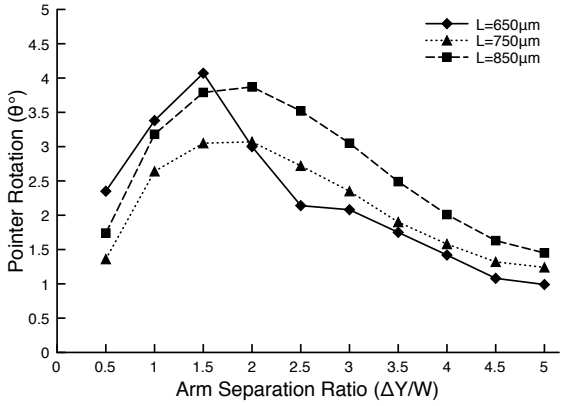



(b) Edge

Fig. 7. Rotation results for 8µm test structures




(a) Middle




(b) Edge

Fig. 8. Rotation results for 10µm test structures

test structures included on the test chip suggest that in most cases the feature sizes are larger near to the middle of the wafer. Assuming that the same is true for the stress sensor structures, it would suggest a higher level of stress is present in the middle of the wafer. However, as this test chip was designed primarily to determine the optimum ΔY/W ratio, this remains unconfirmed.

It should be noted that each data point associated with these results represents a single test structure and the stress may be varying spatially. However, for each test chip on the wafer there are 12 structures of each dimension and further measurements of these devices, to extract their rotations, will supply further information to help determine the variability of the stress measurements.

It was also observed that the pointer arms were deflected downwards at the tip in each of the released test structures. Measurements were made of this deflection using an optical microscope with a split-image rangefinder function to determine the height of features. These measurements, made across the wafer, and similar measurements of the depth of the etched silicon surface, suggested that all of the test structures were bent down until they touched the silicon. Depth measurements taken along a pointer arm suggested that they were all deflected in a smooth arc and that only the very tip of the beam was touching the silicon surface. Calculations of the beam bending under its own weight suggest that this

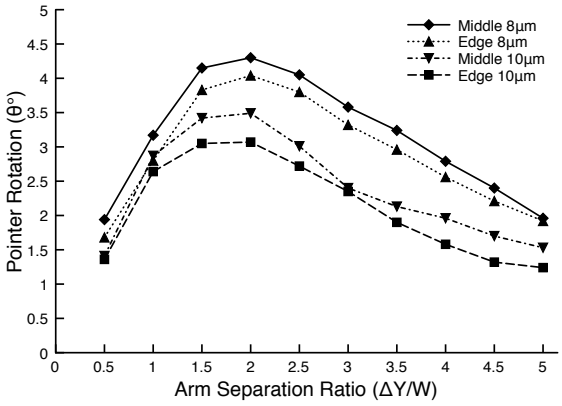


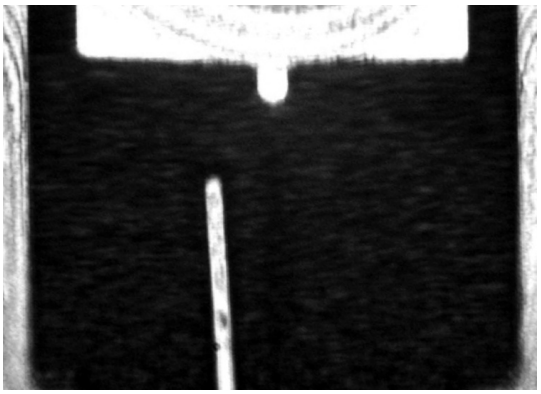

Fig. 9. Comparison of results from structures with L=750µm

would be no more than 200nm [7], which cannot account for the measured deflection. An alternative source for the bending in the pointers is a vertical stress gradient in the permalloy. However, it should be noted that the released structures still have a very thin (<200nm) copper/titanium seed layer on the underside. It is also possible that this could lead to the vertical deflection of the beams due to a mismatch in the residual stress. Unfortunately it is not straightforward to remove this seed layer without damaging the released structures. The fact that the beams appear to be bent down until they touch the surface could have a significant effect on the operation of the test structures, particularly as the silicon etch depth is greater in the middle of the wafer than nearer the edge. Clearly, if some of the tips were stuck and are not free to rotate then this could be the cause of some of the effects observed above.

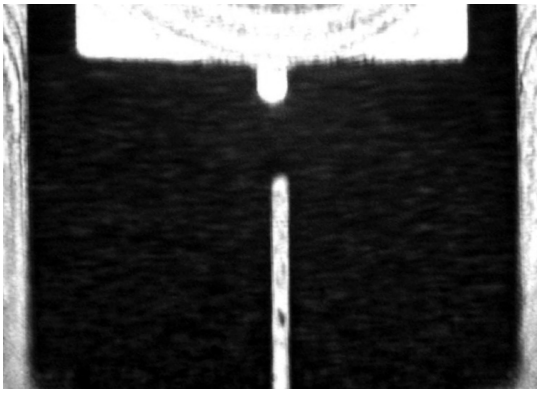
To investigate whether or not the structures are free to move, one of them was heated by passing an electrical current along the expansion arms. The standard stress test structure shown in Fig. 3 is designed to be electrically probed as the two ends of the expansion arms are isolated from each other. A current of 50mA was forced along the structure and was found to cause a significant clockwise rotation of the pointer arm indicating joule heating and thermal expansion. Microscope images of one pointer arm taken during the experiment are presented in Fig. 10.

The observed rotation was approximately 3.5° in the opposite direction to the tensile residual stress. When the current was stopped the pointer returned to its initial position indicating that, in this case, the structure was free to move. The rotation was measured at a number of different currents and the results are presented in Fig. 11.

As might be expected the angle of rotation appears to be roughly proportional to the square of the current, i.e., the power dissipated in the structure. The dotted line in Fig. 11 is a quadratic fit to the data. A number of structures were probed with similar results suggesting that stiction was probably not significantly influencing any of the above results.



(a) No current flow



(b) Current of 50mA

Fig. 10. Microscope images of a sensor pointer arm during electrical test

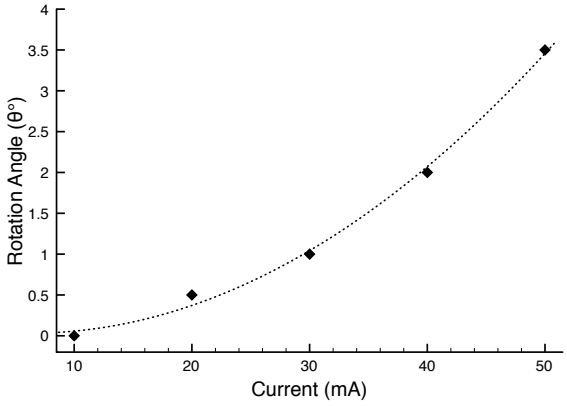



Fig. 11. Pointer arm rotation vs. force current

## CONCLUSIONS

Micromechanical test structures have been used to investigate stress in a relatively thick (>5μm) electroplated permalloy film. A new test chip was designed in order to determine the optimal dimensions for future test structures and results have been presented from the first wafer to be prepared using this design. Anti-clockwise rotations were observed in released sensor structures indicating tensile stress in the deposited layer. The largest rotations were observed on test structures with an expansion arm spacing to width ratio (ΔY/W) of 2. The rotations were generally larger in structures with a narrower expansion arm width, although the narrowest structures did not survive to the end of the process. The rotation, and therefore the sensitivity, was usually larger where the expansion arm length was greater, although there were some unexpected results for the structures with 10μm which have still to be explained. The rotations were slightly higher for structures nearer to the centre of the wafer but there are a number of possible reasons for this apart from a variation of the stress.

It was clear that all of the released pointer arms were deflected downwards with the tips of the arms probably in contact with the surface of the etched silicon. The most likely explanation of this is an unexpected vertical stress gradient in the sensor structures as the deflection cannot be explained as bending of the pointer arms under their own weight. It is possible that some of the measured rotations may not reflect the actual stress in the material if the pointers are not free to move due to contact with the surface. To determine this, measurements were made of a heated stress sensor by passing a current along the expansion arms. This led to a clockwise rotation of the pointer arm which is indicative of Joule heating of the structure causing thermal expansion of the material. The effect of the heating was found to be reversible and the pointer arms returned to their original position when the current was removed. This result demonstrates that this structure is free to move and it is reasonable to assume that the other structures are in a similar condition. More comprehensive measurements are required to confirm this result.

The results presented in this paper demonstrate that these test structures are capable of identifying the presence of compressive and tensile stress in electroplated permalloy. Furthermore they have also identified the design rules that should be used for future test structure designs in order to obtain the greatest sensitivity or the largest rotation for a given stress. Further work will be required to turn this into a quantitative measurement. For example, X-ray diffraction measurements of stress in aluminium films published in [8] suggested that a $1^\circ$ rotation is equivalent to a stress of 28 MPa. However, this is dependent on the material being measured and the dimensions of the structures, and cannot be applied to the results in this paper. A similar calibration will be required in order to extract the actual stress in a permalloy film using these structures. In addition, further investigation into the causes of the vertical deflection of the beams is required. This effect is particularly undesirable if the pointer arm either touches the bottom of the etched trench or is bent down so far that it is difficult to focus on the tip when making measurements of rotation.

One of the objectives of the present work has been to develop a system that can examine the variation of stress across a wafer and this sensor is ideally suited to this purpose. Now that the optimum dimensions have been determined the next test chip will consist of an array of stress test structures to provide characterisation with a high spatial resolution, together with linewidth and Greek cross structures. Cantilever beam structures of different dimensions will also be included to enable characterisation of any beam bending in the vertical direction and help to identify beams touching the silicon surface.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Edinburgh Research Partnership in Engineering and Mathematics. They would also like to thank Memsstar Technology for their help with processing.

## REFERENCES

[1] B. van Drieenhuizen, J. Goosen, P. French, and R. Wolffenbuttel, "Comparison of techniques for measuring both compressive and tensile stress in thin films," *Sensors and Actuators A (Physical)*, vol. 37-38, pp. 756–765, 1993.

[2] X. Zhang, T. Zhang, and Y. Zohar, "Measurement of Residual Stresses in Thin Films Using Micro-Rotating Structures," *Thin Solid Films*, vol. 335, no. 1, pp. 97–105, 1998.

[3] J. Terry, S. Smith, A. Walton, A. Gundlach, J. Stevenson, A. Horsfall, K. Wang, J. dos Santos, S. Soare, N. Wright, A. O. Neill, and S. Bull, "Test Chip for the Development and Evaluation of Sensors for Measuring Stress in Metal Interconnect," *IEEE Transactions on Semiconductor Manufacturing*, vol. 18, no. 2, pp. 225–261, May 2005.

[4] C. Wilson, A. Horsfall, A. O'Neill, N. Wright, S. Bull, J. Terry, J. Stevenson, and A. Walton, "Direct Measurement of Electromigration-Induced Stress in Interconnect Structures," *IEEE Transactions on Device and Materials Reliability*, vol. 7, no. 2, pp. 356–362, June 2005.

[5] M. G. Buehler and W. R. Thurber, "An Experimental Study of Various Cross Sheet Resistor Test Structures," *J. Electrochemical Soc - Solid State Technology*, vol. 125, no. 4, pp. 645–650, April 1978.

[6] M. G. Buehler, S. D. Grant, and W. R. Thurber, "Bridge and van der Pauw Sheet Resistors For Characterizing the Line Width of Conducting Layers," *J. Electrochemical Soc - Solid State Technology*, vol. 125, no. 4, pp. 650–654, April 1978.

[7] N. Wang, "Sheet Resistance, Electrical Linewidth and Micro-rotating Stress Sensor Test Structures for Permalloy Characterisation," MEng Honours Project Report, The University of Edinburgh, 2009.

[8] J. dos Santos, K. Wang, A. Horsfall, J. Pina, N. Wright, A. O'Neill, S. Soare, S. Bull, J. Terry, A. Walton, A. Gundlach, and J. Stevenson, "Calibration of MEMS based test structures for predicting thermo-mechanical stress in integrated circuit interconnect structures," *IEEE Transactions on Device and Materials Reliability*, vol. 5, pp. 713–719, 2005.

# Measurement of MOSFET C-V Curve Variation Using CBCM Method

Katsuhiro Tsuji[1], Kazuo Terada[1], Tomoaki Nakamoto[1], Takaaki Tsunomura[2] and Akio Nishida[2]

Graduate School of Information Sciences, Hiroshima City University[1]
3-4-1 Ozuka-Higashi, Asa-Minami-Ku, Hiroshima, 731-3194, Japan
MIRAI-Selete[2]
16-1 Onogawa, Tsukuba-shi, Ibaraki-ken, 305-8569, Japan

## ABSTRACT

The test circuit, in which the cells including CBCMs (Charge-Based Capacitance Measurements) are arrayed in matrix shape, is developed to measure MOSFET capacitance variation. By adjusting the bias condition of the test circuit, it is able to obtain C-V curves for many MOSFETs. Additionally, a variation of threshold voltage is extracted from the estimated C-V curve variation. The obtained threshold voltage variations are close to those which are obtained from current-voltage characteristics.

## INTRODUCTION

As a MOSFET device dimension is shrunk, variation of device characteristics, such as current and capacitance, has become an important issue. Variation of MOSFET capacitances influences a MOSLSI circuit performance. However, a study of capacitance variation has not been presented so much, while many studies of current characteristics variation have been reported in recent years. The reason is considered that it is difficult to measure the capacitances for many MOSFETs with high accuracy and resolution. Though an LCR meter is usually used to measure a large MOSFET capacitance, it is difficult to measure femto-farad-order capacitance of a small MOSFET due to an influence of parasitic capacitance. CBCM method has been developed to measure interconnect capacitances [1]. It has also been applied to measuring a capacitance of MOSFET, because it can remove parasitic capacitance [2-4]. In this work, to obtain capacitance variation of actual size MOSFET, test circuit is developed using CBCM method. It is structured in DMA (Device Matrix Array) for efficiently measuring a number of MOSFETs [5]. In addition to measuring variation of MOSFET capacitance-voltage (C-V) curve, an extraction of threshold voltage variation is attempted.

## CBCM TEST STRUCTURE

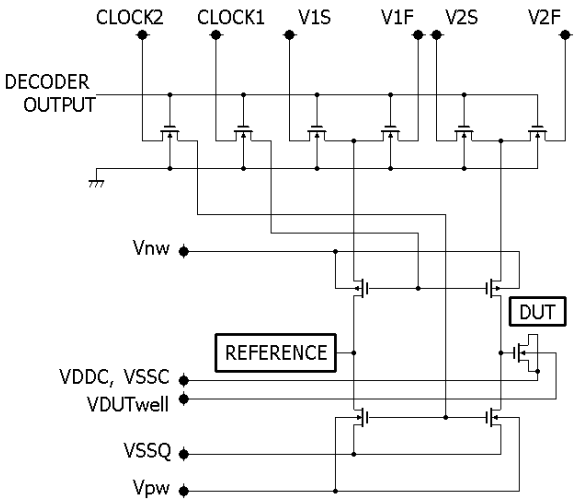



Fig. 1. Unit cell structure of test circuit using CBCM method.

The test circuit contains 64cells, decoders and common probing pads. Each cell consists of NMOS transfer gates, CBCM and DUT (Device Under Test) MOSFET. All MOSFETs used in the circuits other than CBCM and DUT, which are decoders and NMOS transfer gates, are designed with 0.6-μm channel and 3-V supply technology. Fig. 1 shows the cell structure of the test circuit other than decoder output circuit. It consists of a pair of pseudo-inverter that NMOS and PMOS are connected in series. A DUT is connected to middle point of both NMOS and PMOS in pseudo-inverter. DUT MOSFET and pseudo-inverter are designed with 65-nm channel and 1.2-V supply technology. To increase measurement accuracy, Kelvin connection is employed at terminals that are used to measure charge and discharge current. The last characters F and S in terminal name denote "Force" terminal and "Sense" terminal for Kelvin connection, respectively. When this cell is selected, nodes V1F, V1S, V2F, V2S, CLOCK1 and CLOCK2 are connected to

common probing pads through NMOS transfer gates. This test circuit is also featured by that every voltage supply terminals of pseudo-inverter and DUT MOSFET can be controlled independently. At gate terminals of PMOS and NMOS in pseudo-inverter, two non-overlapping signals with same frequency are provided through CLOCK1 and CLOCK2 terminals from an external pulse generator. Capacitance of DUT $C_{DUT}$ is obtained by measuring charge-discharge current flowing at V1F and V2F terminals from the following expression:

$$C_{DUT} = \frac{I_{DUT} - I_{REF}}{f(VDDQ - VSSQ)} \quad , \qquad (1)$$

where $I_{DUT}$ and $I_{REF}$ are measured currents at V2F terminal and at V1F terminal, respectively. f is a frequency of clock input signal, and VDDQ is an applied voltage of both V1F and V2F. VSSQ is NMOS source voltage in pseudo-inverter. A measured current by using CBCM method is an average value between charge voltage VDDQ and discharge voltage VSSQ. Therefore, it is considered that an obtained $C_{DUT}$ from (1) is also an average capacitance. If parasitic components of both pseudo-inverters are identical, in the case of cell without DUT to one side, $C_{DUT}$ is simply estimated from (1). However, since each pseudo-inverter has a variation, we need to take account of it to estimate DUT capacitance. In this paper, because DUT dimension is enough large as compared with MOSFETs dimensions in pseudo-inverter, an average current that is obtained from several cells without DUT is used to $I_{REF}$.

## RESULTS of CAPACITANCE MEASUREMENT

Measurements of charge and discharge current are performed by semiconductor device analyzer B1500A. Fig. 2 shows the plot of capacitances versus VSSQ voltage for 16DUTs. The measured DUT dimension is channel width W = 1μm and channel length L = 1μm. The input clock frequency is 1MHz with voltage swing from 0 to 1V. In the measured cells, DUT MOSFET of same size is connected to not only the DUT node but also the REFERENCE node in Fig. 1. Therefore, because it is not able to simply subtract the current flowing at V1F terminal from that at V2F terminal, the obtained capacitances in Fig. 2 include parasitic component other than DUT MOSFET. As VSSQ increases, though the capacitances increase gradually, they rapidly decrease at VSSQ > 0.5. It is because NMOS of pseudo-inverter turns off and applied voltage on VDDQ is constant at 1V. The turn-off points are different among 16DUTs, because of threshold voltage variation of MOSFET in pseudo-inverter. When VDDQ voltage is changed and VSSQ voltage is constant, similar characteristics are observed due to the influence of PMOS threshold voltage.

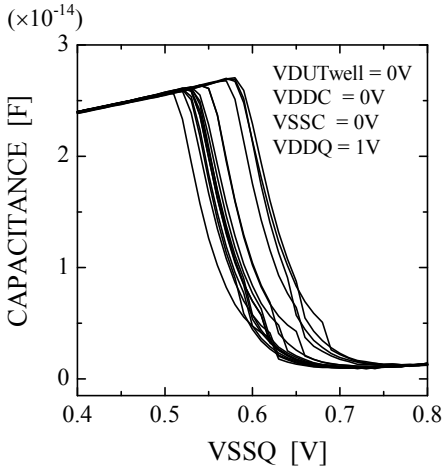



Fig. 2. Plot of capacitances versus VSSQ voltage.

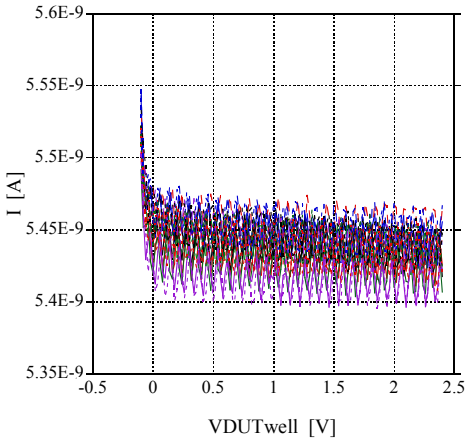



Fig. 3. Current characteristics without DUT, when VDUTwell terminal voltage is changed.

Therefore, to obtain continuous MOSFET C-V curve including accumulation, depletion and inversion condition, a bias condition of current measurement has to be implemented, as avoiding the influence of threshold voltage of MOSFET in pseudo-inverter. Thus, we perform a measurement of changing VDUTwell voltage in Fig. 1.

Fig. 3 shows current characteristics changing VDUTwell terminal voltage without DUT, when VDDQ and VSSQ are fixed at 1V and 0.5V, respectively. VDDC and VSSC are also equaled to a voltage of VDUTwell during a measurement. It is found that oscillation is observed in the 16 measured current characteristics. The each range of oscillation is also around 50pA. The cause of oscillation may be mainly attributed to noise from power

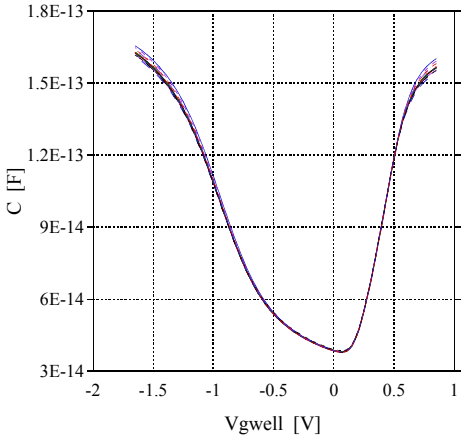



Fig. 4. Capacitance versus voltage characteristics for 16DUTs (W/L = 4μm/4μm).

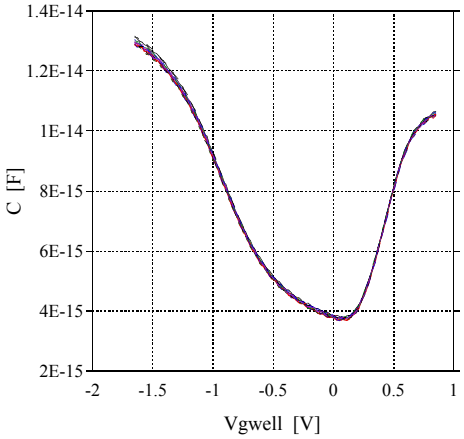



Fig. 6. Capacitance versus voltage characteristics for 16DUTs (W/L = 1μm/1μm)

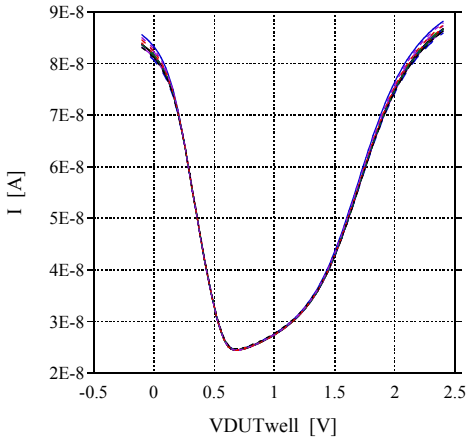



Fig. 5. Measured current for 16DUTs (W/L = 4μm/4μm), when VDUTwell is changed.

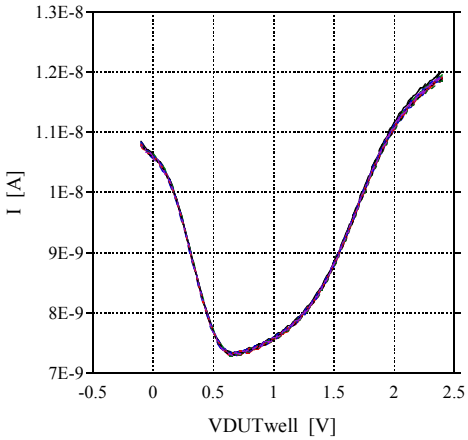



Fig. 7. Measured current for 16DUTs (W/L = 1mm/1mm), when VDUTwell is changed.

supply and stray capacitance of peripheral circuit. Furthermore, the slight difference between each characteristic is caused by the variation of the source/drain junction and overlap capacitances between the pseudo-inverter MOSFETs within different cells. In order to remove the influence of them, the measured current is approximated by averaging procedure at every measurement voltage. The obtained average data from the above-mentioned way is used as $I_{REF}$ to estimated DUT capacitance by (1).

Fig. 4 shows the estimated C-V curves for 16DUT MOSFETs with W/L = 4μm/4μm. These plots are obtained from the measurement of changing VDUTwell voltage which is pwell voltage of DUT MOSFET. During the measurement, terminal voltages other than VDUTwell are given as follows: VDDQ = Vnw = 1V, VSSQ = Vpw = 0.5V and VDDC = VSSC = VDUTwell. Since VDUTwell voltage is changed, the voltage is converted to the effective gate-substrate voltage of DUT MOSFET by using (2).

$$\mathrm{Vgwell} = (\mathrm{VDDQ} + \mathrm{VSSQ})/2 - \mathrm{VDUTwell} \qquad (2)$$

Therefore, it is considered that these capacitance characteristics represent accumulation, depletion and inversion conditions from left side toward right side in Fig. 4. The influences of oscillations existing in Fig. 3 and MOSFETs variation of pseudo-inverter are not observed due to large DUT dimension. The reason is considered to be that the measured current with DUT is due to enough large as compared with the current without DUT (see Fig. 3) as shown in Fig. 5. As a result, it is considered that C-V characteristics of DUT MOSFET are obtained.

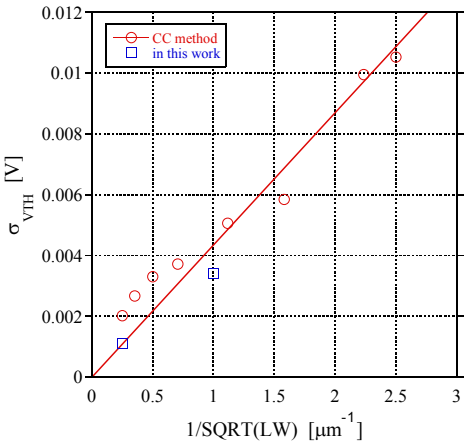



Fig. 8. Comparison of threshold voltage variations by Pelgrom Plot.

Fig. 6 shows the estimated C-V curves for 16DUTs with W/L = 1μm/1μm. The bias condition of measurement setup is identical to the case of Fig. 4. Since DUT dimension becomes small, the influence of them observed in Fig. 3 is slightly appeared. Fig. 7 shows the measured current for 16DUTs with W/L = 1μm/1μm. Comparing these characteristics with those of Fig. 3, it is found that the difference of current value is little. Consequently, it is considered that the variation of MOSFETs in pseudo-inverter is slightly appeared. However, it is concluded that the obtained results by this method are regarded as MOSFET C-V curve, because the influence has little effect on the stability of estimation.

### Comparison of variation

From above-mentioned results, we extract the variation of MOSFET threshold voltage $\sigma_{VTH}$. Assuming that the estimated C-V curve has appropriate accuracy, a variation of device parameter, such as threshold voltage, can be obtained from the variation of C-V curves. To extract $\sigma_{VTH}$ from C-V characteristics, threshold voltage is defined as a voltage at a given constant capacitance. The results of the $\sigma_{VTH}$ extraction are shown in Fig. 8. The squares denote the obtained variations from above-mentioned results. The circles show the extracted $\sigma_{VTH}$ by constant current (CC) method [6]. The current voltage characteristics are measured from another chip which was fabricated by same technology. The straight line in Fig. 8 is drawn by Pelgrom Plot [7]. The plot has been commonly used to evaluate a variation of MOSFET parameter. As compared the obtained $\sigma_{VTH}$ in this work to Pelgrom Plot, it is found that they take a value near the straight line of variation by CC method. Consequently, it is considered that the developed test circuit is useful to extract a variation of MOSFET threshold voltage. However, the obtained variations by this work slightly include the influence of variation of the junction and overlap capacitances in the MOSFETs which configure pseudo-inverter, because of averaging procedure of reference current for several cells.

## CONCLUSION

DMA-like test circuit is developed to measure MOSFET capacitance by using CBCM method. The developed circuit is able to be simply measured a lot of MOSFET capacitances with appropriate accuracy. It is found that the obtained capacitance characteristics can be regarded as MOSFET C-V curve. Furthermore, the estimated variation of C-V curves is useful to evaluate variation of device characteristics.

## ACKNOWLEDGMENT

A part of this work is supported by NEDO.

## REFERENCES

[1] J. C. Chen, B. W. McGaughy, D. Sylvester, and C. Hu, “An on-chip, attofarad interconnect charge-based capacitance measurement (CBCM) technique,” in *IEDM Tech. Dig.*, 1996, pp. 69-72.

[2] B. Sell, A. Avellan, and W. H. Krautschneider, “Charge-based capacitance measurement (CBCM) on MOS devices,” IEEE Trans. Device and Materials Reliability, vol. 2, no. 1, pp. 9-12, March 2002.

[3] T. Sutory, and Z. Kolka, “Characterization of nonlinear on-chip capacitors,” in *Proc. 17th Intl. Conf. Radioelektronika*, April, 2007, pp. 51-54.

[4] Y. Chang, H. Chang, T. Lu, Y. King, K. Chen, and C. Lu, “Combining a novel charge-based capacitance measurement (CBCM) technique and split C-V method to specifically characterize the STI stress effect along the width direction of MOSFET devices,” IEEE Electron Device Lett., vol. 29, no. 6, pp. 641-644, June 2008.

[5] S. Ohkawa, M. Aoki, and H. Masuda, “Analysis and characterization of device variations in an LSI chip using an integrated device matrix array,” in Proc. ICMTS2003, 2003, pp. 70-75.

[6] H. Lee, S. Oh, and G. Fuller, “A simple and accurate method to measure the threshold voltage of an enhancement-mode MOSFET,” IEEE Trans. Electron Devices, vol. ED-29, no. 2, February, 1982.

[7] M. J. M. Pelgrom, H. P. Tuinhout, and M. Vertregt, “Transistor matching in analog CMOS applications,” *IEDM Tech. Dig.*, pp. 915 - 918, December 1998.

# Array Test Structure for Ultra-Thin Gate Oxide Degradation Issues

Kristian M. Hafkemeyer, Andreas Domdey, Dietmar Schroeder and Wolfgang H. Krautschneider

Institute of Nanoelectronics, Hamburg University of Technology, Germany, D-21073 Hamburg
Phone: +49 40 42878 2919, Fax: +49 40 42878 2877, Email: kristian.hafkemeyer@tu-harburg.de

**ABSTRACT**

An array test structure for highly parallelized measurements of ultra-thin MOS gate oxide failures caused by degradation is presented. The test structure allows for voltage stress tests of several thousand NMOS devices under test (DUTs) in parallel to provide a large and significant statistical base regarding soft as well as hard breakdown and stress induced degradation of transistor parameters. The array has been fabricated in a standard 130 nm CMOS technology. As mixed mode technologies provide both thin and thick oxide MOS transistors, different gate oxide thicknesses have been chosen for DUTs and digital control logic which gives the possibility to stress the DUTs with high gate voltages.

## I. INTRODUCTION

Within the past decade the minimum feature size of MOSFETs has been scaled dramatically [1]. Oxides in CMOS circuits have reached 1-2 nm in thickness [2], [3] that is accompanied by increasing reliability constraints [4]-[7]. The high gate current density of such small devices may effectuate a continuous degradation of the gate oxide until malfunction of the device in terms of soft breakdown (SBD) or hard breakdown (HBD) occurs. The HBD is defined as an abrupt increase in gate tunneling current by several orders of magnitude [8]. In the post breakdown regime, the devices with SBD continue to degrade until HBD [9]. SBD is identified as a sudden increase in gate current from an initial baseline current by at least one order of magnitude [10]. A large statistic about the degradation process as a function of time gives insights into transistor degradation and parameter variation. Array test structures already exist to investigate the MOSFET properties on chips or wafers [11], [12]. In addition, the test structure presented here allows for monitoring the gate current for each DUT individually.

The possibility to measure the gate tunneling current during constant voltage and temperature stress speeds up the measurement tremendously, since single rows can be disconnected from voltage stress while keeping the array under stress. This enables the early detection of SBDs or HBDs and the deactivation of broken-down DUTs from subsequent stress. Accelerated tests require high gate voltages and increased temperatures. To accommodate this requirement, the two different gate oxide thicknesses of the used 130 nm CMOS mixed-mode technology are utilized. For the MOS DUTs, an oxide thickness of 2.2 nm and for the periphery (cell logic, decoders) 6.5 nm is used, respectively. The periphery can switch high stress voltages for the thin oxide devices without degradation.

In this paper, an integrated array structure is proposed for the evaluation of transistor parameters during voltage and temperature stress. Furthermore, statistics of the quality on wafer level by means of process gradients can be ascertained.

At first, the array structure itself and the measurement setup are introduced in Sec. II. Results from the measurements are presented in Sec. III. Conclusions are finally given in Sec. IV.

## II. TEST STRUCTURE

### *A. System overview*

Fig. 1 shows an improved structure of the first generation [13] of the integrated test array. The matrix-like arrangement of the DUTs allows for stress tests of $2^N$ x $2^N$ DUT cells in parallel under same conditions. For the presented test array, we have chosen N = 5 for 1024 DUTs.

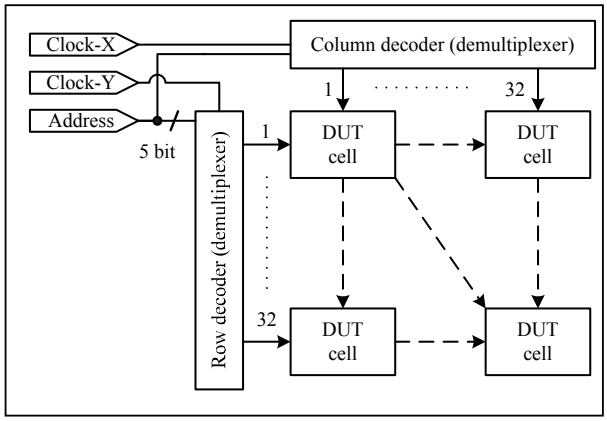



Fig. 1. Illustration of DUT cell array test structure. 1024 DUTs can be stressed in parallel. Each device can be measured by selecting its address.

Individual DUTs can be accessed using demultiplexing decoder circuits. A shared address input for the row and column decoders reduces the number of pads on chip. Each decoder is triggered with a clock signal applied to its input. In this case, the output registers of the decoders are set to the appropriate value.

The thick oxides can be exposed to high gate voltages much larger than 3.3 V. This option has been utilized to stress the DUTs nearly up to the supply voltage of the thick oxide transistors.

As there were no digital libraries with thick oxides for this technology available, a library with digital standard gates consisting of 6.5 nm thick oxide transistors had to be designed. This includes simple NAND, NOR and INV gates as well as D-type flip-flops.

*B. Single DUT cells*

Fig. 2 shows the DUT cell circuit in combination with a row switch. Each cell consists of digital control logic and an NMOS device under test. The “reset”, “data”, “clock” and “select all” inputs are connected to each cell in the array. The cells can be accessed individually with the AND-gate combining the row and column input. “Select all” overrides the row and column input through the OR-gate. “Data” is stored in the D-type flip-flop of the selected cell at positive clock edge. Depending on the state of the flip-flop, a transfer transistor connects the DUT gate to a supply line common to all cells in its row (SStress) or sets the gate with a discharging transistor to ground. The possibility to toggle the gate stress voltage facilitates the deactivation of devices suffered from HBD as such devices leave a conducting path inside the gate. Thus, they may obscure the remaining gate tunneling currents in a row. For the sake of simplification a common drain (CDrain) and source (CSource) are used. Nevertheless, IV-characteristics of the single DUT can be analyzed when all other elements of the array are deactivated. A switch in front of every row separates the common stress (CStress) of the matrix from the measurement of single addressed rows (SStress). The idea behind that is to stress the matrix while measuring single DUTs.

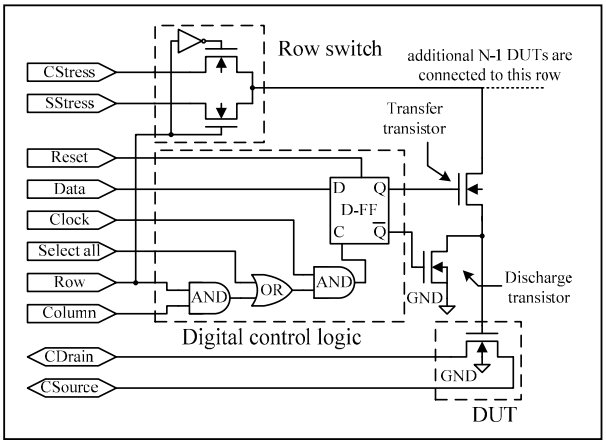



Fig. 2. Row switch and digital control logic of every DUT cell enables individual access to the single MOS device. High resolution of measurements is ensured by ability to connect every line of DUT cells to external ampere-meter separately.

The leakage current through the deselected transfer transistor and the other row switches are of great importance. To distinguish the tunneling current from leakage current, the former has to be in the range of or even larger than the leakage current. Since the discharge transistor is inversely active to the transfer transistor, the gate of the DUT is always grounded if the cell is deselected. Therefore, the static leakage current through the transfer transistor is rather independent, no matter whether a DUT suffers from HBD or not. This increases the reliability in the decision of SBDs. The static leakage current can be further attenuated if a triple well option is used for the pass-gates.

The dependence on temperature of the tunneling current through the gates and the leakage current of the transfer transistors have been investigated. The gate current in tests with enhanced temperature is distinguishable from the leakage current up to 50 °C. The temperature range can be extended if the measurement voltage is increased by approx. 230 mV per 10 Kelvin in worst case.

*C. Design of the integrated test structure*

Fig. 3a shows the layout of the array structure and Fig. 3b of a single cell respectively. The DUT is surrounded by a substrate contact. Wide metal has been used for CStress, SStress, CDrain and CSource to prevent from electromigration caused by high currents through HBD. Minimization of parasitic resistances is another positive effect. Each DUT cell has a size of 16x21 µm² and is large due to the digital control logic compared to the DUT itself. The whole array size of 600x770 µm² includes the decoders and 1024 DUT cells.

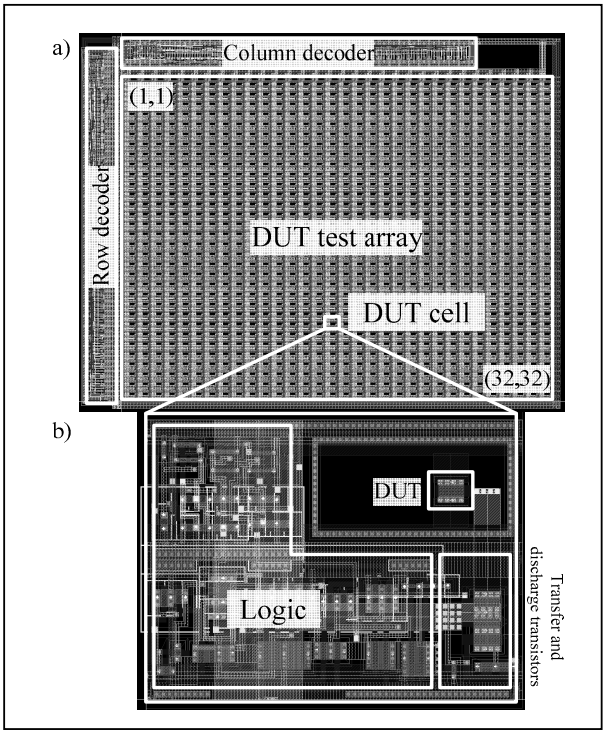



Fig. 3. Layout of the DUT test array and single cell fabricated in a mixed-mode 130 nm CMOS technology.

*D. Measurement Setup*

A parametric analyzer (Keithley 4200-SCS) as shown in Fig. 4, or alternatively, an automated test system presented in [14], is connected to the matrix for high precision measurements with picoampere resolution.

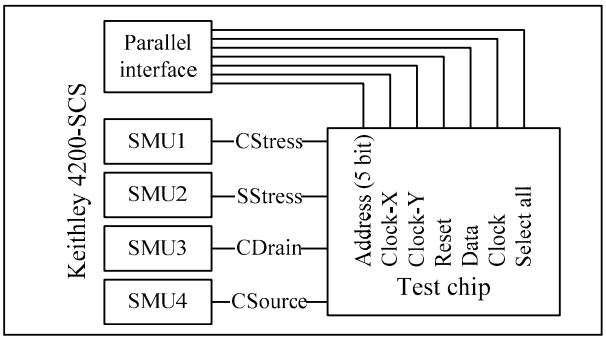



Fig. 4. Semiconductor Characterization System Keithley 4200 SCS is used for controlling and biasing the test structure and for acquisition of measurement data.

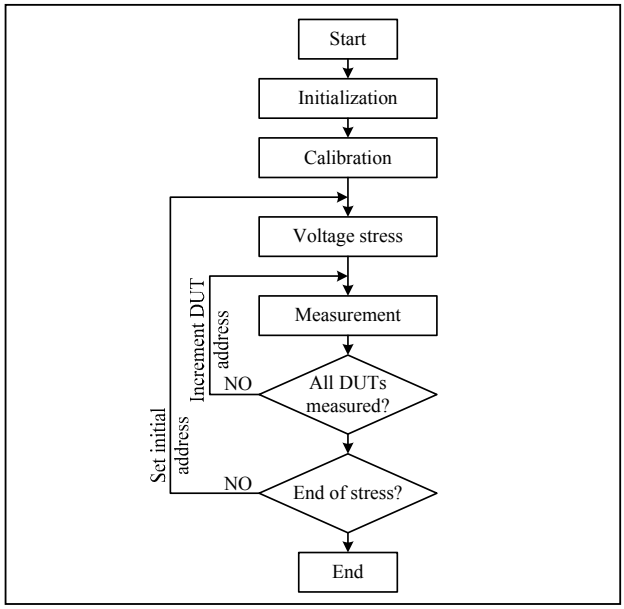



Fig. 5. Flow-chart of the software for controlling the test chip. The measurement cycle is repeated until end of the stress.

It runs the software for controlling the array structure and the data acquisition of the gate current of addressed devices presented in Fig. 5.

At first, the initialization step deselects all DUTs preliminary to the calibration. During the calibration all static leakage currents are measured in each row. The voltage stress begins with the connection of all DUTs to the stress voltage "CStress". After a time limited constant voltage stress, all devices are deactivated and every DUT cell is measured individually through the supply line "SStress". Subsequent to the voltage stress, the address pointer is incremented. This procedure is repeated until all DUTs have been measured and the voltage stress is finished. The IV characteristics of each DUT cell can be measured on demand.

## III. Experimental Results

At the beginning of the stress test, the structure is characterized regarding its leakage currents through each single row in order to calibrate the measurement. Fig. 6 shows the measured leakage current of the test array at room temperature. A mean current of approximately 4 pA flows through an off-switched DUT cell, which agrees with simulation data. The leakage current of 31 deselected DUT cells is considered in the calculations for the SBD detection. The gate length of the NMOS devices under test has been

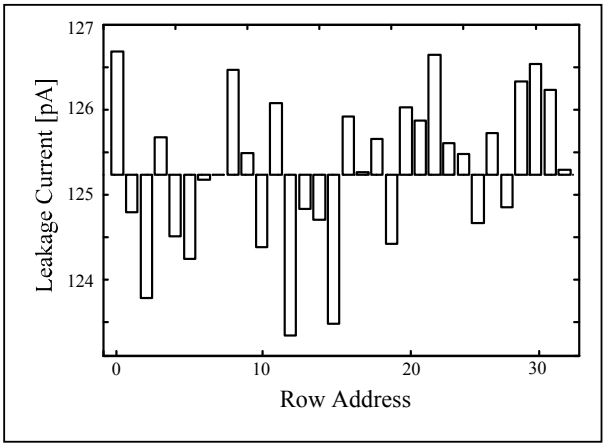



Fig. 6. Leakage current through each row of the test structure. The mean value of the leakage current is 125.26 pA with a max. deviation of 1.5 %.

chosen to L = 0.3 µm and a width of W = 1.9 µm, respectively. Since the subthreshold current of the transfer transistor is strongly dependent on the temperature this calibration step is mandatory.

To illustrate the variation and time evolution of typical breakdown behavior, exemplary devices were taken as depicted in Fig. 7. From measurement data it was found out that a threshold gate current for an SBD event of one order of magnitude related to the base-line current as defined in [10] is too high for detecting DUTs with SBD only. Therefore, the definition for SBD has been changed as at least 10 % increase in gate current from the initial base-line current in order to distinguish between direct HBD and HBD preceding SBD.

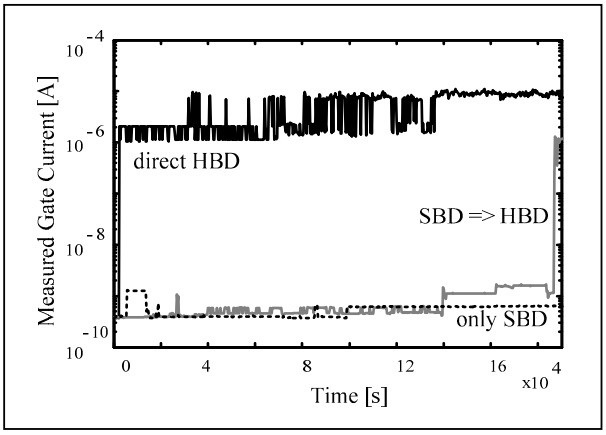



Fig. 7. Observed types of typical breakdown behavior ($V_{GS}$ = 1.8 V). DUTs with L = 0.3 µm and W = 1.9 µm have been stressed.

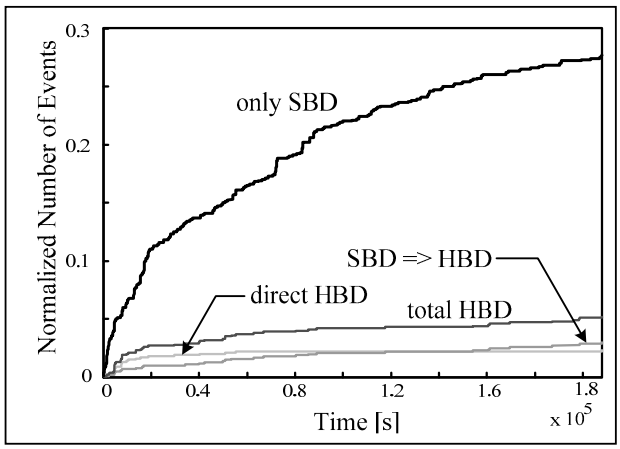



Fig. 8. Cumulative occurrence of breakdown events normalized to the total number of DUTs.

The breakdown history of 1024 DUTs during voltage stress of 200 ksec is shown in Fig. 8. The breakdown behavior obtained from measurements, shown by approximately 30 % of the DUTs, can be classified into three types. Several DUTs exhibit SBD (only SBD). Another small number of DUTs suffers HBD (SBD=>HBD) after SBD. The third type of DUTs shows direct HBD without preceding SBD. The number of direct HBDs levels off after a few stress cycles, representing early failures. All other devices exhibit no remarkable behavior.

Fig. 9 indicates the gate current density distribution by position in a test array after stress. The upper part of the figure shows the gate current density of the devices without HBD.

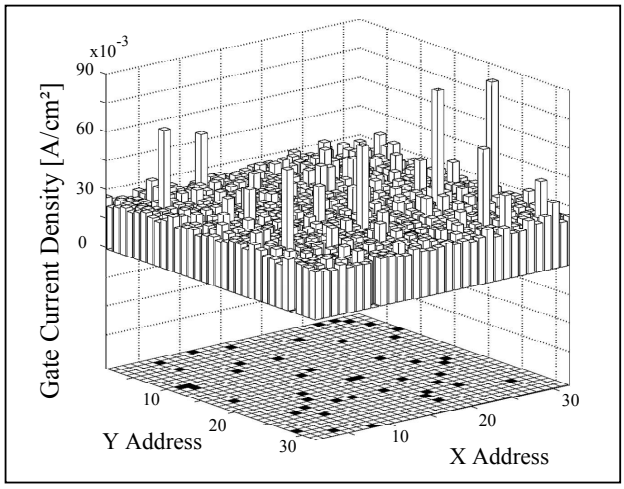



Fig. 9. Gate current distribution by position in the array ($V_{GS}$ = 1.8 V, $V_{DS}$ = 0.0 V).

Peaks with 10 % increase in current density represent devices with SBD (27 %), located over a noisy floor of unbroken devices (68 %). The lower part of the figure shows the positions of devices with HBD (2.1 % direct HBD, 2.8 % SBD => HBD events).

Fig. 10 shows the IV-transfer characteristics as well as the standard deviation of the current value of NMOS devices under test obtained from the array structure. The threshold voltage $V_{TH}$ is extracted from measured data using the linear extrapolation method [15]. The mean threshold voltage at the beginning of the stress cycle is calculated to $V_{TH}$ = 0.34 V and follows a Gaussian distribution having a standard deviation of $\sigma$ = 4 mV over the sample array. The threshold voltage and the standard deviation shift to $V_{TH}$ = 0.38 V and to $\sigma$ = 10 mV, respectively, after stress for elements which exhibited only SBD or are unbroken. The threshold voltage and the standard deviation of the stressed DUTs have increased. Note that the transfer characteristic and thus the threshold voltage is the same in unbroken devices and devices with SBD.

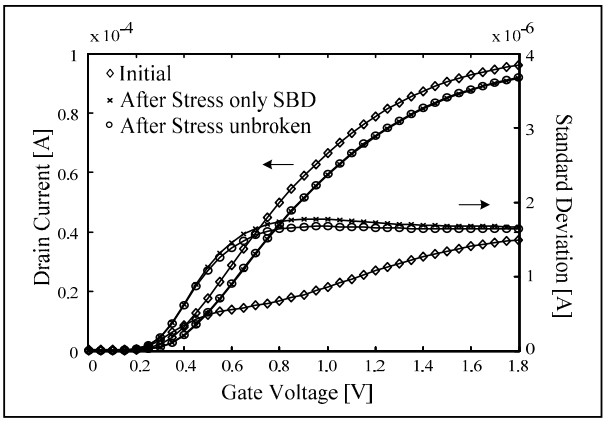



Fig. 10. Mean IV-transfer characteristic as well as standard deviation of NMOS transistors in test array ($V_{DS}$ = 0.05 V) before and after stress which are unbroken or suffered only SBD.

## IV. Conclusion

The array test structure gives detailed insight into the breakdown behavior of thousands identical MOS DUTs that can be stressed and measured in parallel with high gate voltage even at increased temperatures. Thus, it is possible to focus on the relatively small percentages of breakdown events. The correlation between SBD and HBD can be studied.

Furthermore, IV-characteristics can be measured at any time interrupting the stress. Therefore, the shift of the threshold voltage and the standard deviation of the drain current degradation can be determined regarding the position of the device in the array and on wafers. All of this shows the variability in use of the proposed test structure.

## Acknowledgement

This research work is sponsored by the German research foundation "Deutsche Forschungsgemeinschaft (DFG)", Contract No. KR 3281/1-1.

## References

[1] The International Technology Roadmap for Semiconductor, Semiconductor Industry Association, 2000, 2007.

[2] R. Chau, S. Datta, M. Doczy, J. Kavalieros and M. Metz, "Gate Dielectric Scaling for High-Performance CMOS: from SiO2 to High-K", IWGI, p. 124, 2003.

[3] P. Bai, S. Balakrishnan, M. Bost et al, "A 65nm Logic Technology Featuring 32nm Gate Length, Enhanced Channel Strain, 8 Cu Interconnect Layers, Low-k ILD and 0.67µm² SRAM cell", IEDM Technical Digest, p. 657, 2004.

[4] B.E. Weir, P.J. Silverman, D. Monroe et al, "Ultra-thin gate dielectrics: they break down, but do they fail?", IEDM Technical Digest, p. 73, 1997.

[5] B. Kaczer, R. Degraeve, M. Rasras, K. Van de Mieroop, P. J. Roussel, and G. Groeseneken, "Impact of MOSFET gate oxide breakdown on digital circuit operation and reliability", IEEE Trans. Electron Devices, Vol. 49, p. 500, 2002.

[6] A. Avellan and W.H. Krautschneider, "Impact of Soft and Hard Breakdown on Analog and Digital Circuits", IEEE Transactions on Device and Materials Reliability, Vol. 4, No. 4, p. 1, 2004.

[7] R. Doering and Y. Nishi, Handbook of Semiconductor Manufacturing Technology, CRC Press, 2008.

[8] K. Okada, "Extended time dependent dielectric breakdown model based on anomalous gate area dependence of lifetime in ultra thin silicon dioxides", Jpn. J. Appl. Phys., Vol. 36, p. 1434, 1997.

[9] F. Monsieur, E. Vincent, D. Roy et al, "A Thorough Investigation of Progressive Breakdown in Ultra-Thin Oxides. Physical Understanding and Application for Industrial Reliability Assessment", Proc. of IRPS, p. 45, 2002.

[10] S. Lee, B. Cho, J. Kim, and S. Choi, "Quasibreakdown of ultrathin gate oxide under high field stress", IEDM Tech. Dig., p. 605, 1994.

[11] Y. Shimizu, M. Nakamura, T. Matsuoka et al, “Test Structure for Precise Statistical Characteristics Measurement of MOSFETs“, Proc. of ICMTS, Vol. 15, p. 49, 2002.
[12] N. Izumi, H. Ozaki, Y. Nakagawa et al, “Evaluation of Transistor Property Variations Within Chips on 300-mm Wafers Using a New MOSFET Array Test Structure”, IEEE Trans. Semi. Cond. Manufacturing, Vol. 17, No. 3. p. 248, 2004.
[13] K. Hafkemeyer, A. Domdey and W.H. Krautschneider, “Design of a process parameter independent test structure for reliability tests”, ProRISC, p. 266, 2006.
[14] A. Domdey, K. Hafkemeyer, D. Schroeder and W.H. Krautschneider, “Automated system for simultaneous large-scale stress tests of ultra-thin gate dielectrics to analyse reliability”, ProRISC, p. 14, 2007.
[15] L. Dobrescu, M. Petrov, D. Dobrescu et al, “Threshold Voltage Extraction Methods for MOS Transistors”, Proc. of CAS, Vol. 1, p. 371, 2000.

# Non-Contact, Pad-less Measurement Technology and Test Structures for Characterization of Cross-Wafer and In-Die Product Variability

Gary Steinbrueck, James S. Vickers, Majid Babazadeh,
Mario M. Pelella, and Nader Pakdaman
*tau-Metrix Inc. 2350 Mission College Blvd.*
*Santa Clara, CA 94054 (*gs@tau-Metrix.com*)*

**Abstract**

Monitoring and controlling cross-wafer and in-die variability has been recognized as the dominant and escalating factors for the successful commercialization of modern-day integrated circuit products utilizing advanced semiconductor manufacturing[1,2]. In this paper we present a Performance Based Metrology (PBM), a measurement technology for closing the information gap between the design, process integration, and manufacturing groups with respect to accounting for variability. PBM enables the "porting" of scribe-like and end-of-line contact tested measurements to within the product die active area to it provide the capability to significantly reduce the cycles of learning to obtain relevant and key process, device, and product metrics. This product-relevant information can then be used for process monitoring and control, performance optimization, and to enhance early bin-yield predictability. The technique would reduce or eliminate the need for send-ahead test wafers and other "disruptive" measurements by making possible in-die, non-contact characterization of product performance monitors and devices. We describe the in-line measurement system and review the design and implementation considerations for the non-contact test structures incorporated on product wafers. Experimental results from PBM measurements on several generations (90, 65, and 45nm) of bulk-Si and SOI product wafers and devices are presented to illustrate the capabilities of the technique.

## Introduction

The increasing cost and complexity of semiconductor processes and designs is creating new demands for timely and continuous feedback on performance between the groups responsible for controlling the cost and assuring full functionality of their products. A key determinant of profitability has become how and when the variability inherent in the manufacturing processes and the sensitivity of ever more complex products to that variability gets reconciled. Monitoring results and developing useful information on which to base adjustments and corrective action now depends on more direct and statistically valid measurement of product performance, with greater resolution, and earlier in production than ever before.

The value of information about process induced performance variation can be gauged by the cost, time, and effort being expended on methods to estimate and account for the effects of it. These include incorporating advanced statistics into the design process, defining hundreds of rules for each mask layer, making adjustments to the mask and wafer images, using finer process control with more sophisticated materials and structures, building in functional and physical redundancy, integrating test structures into the product, and more. All of these, represented by a mountain of acronyms and many millions of dollars, have become necessary and will continue to increase in importance due to the effects of variation on the performance of integrated circuits. Such methods will yield maximum benefit when the rules for their application and feedback on their effectiveness are derived directly from a product's active devices as early as possible in the development cycle and during the manufacturing process.

The high price of progress has driven economic adaptations as well: from integration and consolidation of manufacturing facilities, to distribution and fragmentation of product development, and even the balance of risks and benefits from decisions affecting product time-to-market and functionality. Every interface between parties, be they tool managers or companies, having responsibility for the success of a product implies a contract, whether formal or informal. In the evolving world of semiconductors, enforcible contracts with measurable objectives have become the key to successful operation. There was a time when integrated device manufacturers could gain some advantage by adjusting their process to their products, today these production lines are often built and managed by multi-company alliances, filled with many types of products designed world-wide; all having different sensitivities to the process, and all under contract. Where once a fab-less enterprise could guard against unknowns in the manufacture of their products by designing conservatively or accepting yield losses, the tangible and intangible costs of this are now becoming prohibitive. Variation must become more of a known quantity to get maximum return from every design and every line.

978-1-4244-4259-1/09/$25.00 ©2009 IEEE 

PBM, Performance Based Metrology, allows the measurement of in-die variation on product-like test structures at the early stages of manufacturing when first circuit connectivity is realized (typically first metal). This "early" information can be used to monitor and control performance variation for line-to-line (L2L), run-to-run (R2R), wafer-to-wafer (W2W), and die-to-die (D2D) design/ product/process monitor and control. The need to complement and manage traditional and still critical metrology techniques with parametric in-line, in-die measurement which correlates directly to product performance and ultimately final bin-yield has been recognized[3]. PBM accelerates the gathering and availability of the current critical measurements of product representative[4] RO-type test structures in the scribe and die[5] from early stages of product development and continuously throughout the manufacturing process steps and phases. This is achieved with laser-stimulated, pad-less power/sense targets that enable the non-contact measurement of the product-like ROs in the active product die and scribe.

### Performance Based Metrology (PBM) Architecture

The PBM system encompasses two functional blocks: 1) the non-contact power delivery and measurement component; 2) the embedded (in product silicon) structures to enable the non-contact activation (power) and measurement of the product-representative test structures.

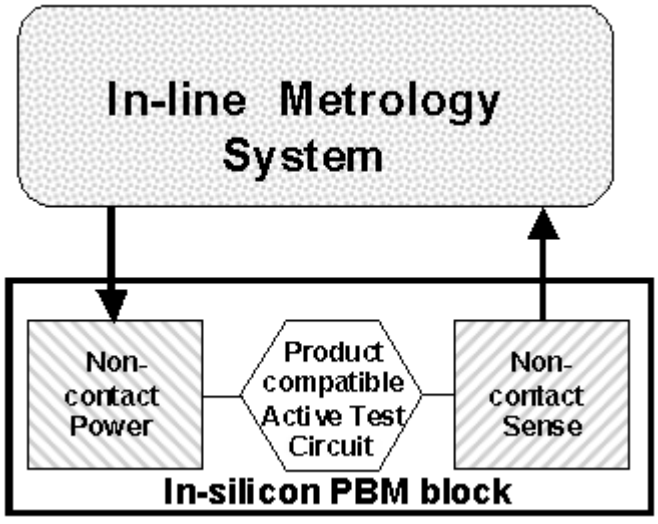



Fig. 1 PBM active test target and measurement system

The overall system architecture is shown in Fig. 1. The in-line metrology system comprises an automated wafer movement and handling system, along with the associated imaging system to enable fast alignment and positioning. The power delivery and coupling to the power-structure on the silicon and the signal detection rely on optoelectronic effects and devices. This equipment is fully compatible with advanced manufacturing systems.

### Pad-less Non-Contact Power and Signal Detection Design and Integration To ROs

The active test structures, fabricated in the silicon wafer, are comprised of ring oscillators and power/sense devices. An example is shown in Fig. 2:

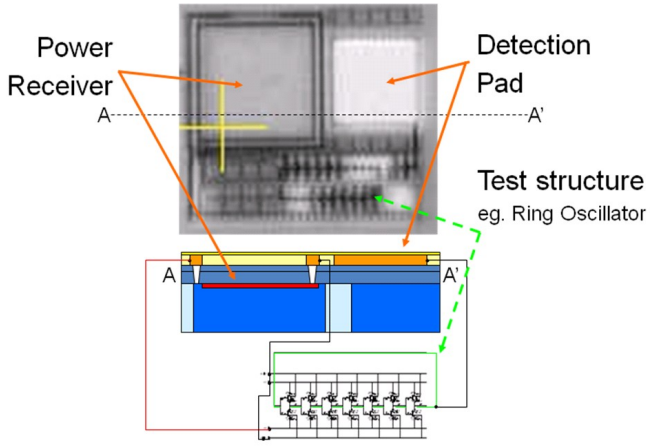



Fig. 2 Example in-silicon PBM block

These elements are formed with standard product mask layers and process steps - no additions and/or modifications to the standard product definition and associated fabrication process are needed. Integration into product or scribe requires adherence to standard design rules (and, if desired, definition of device recognition for compatibility with production design checkers). Non-contact signal sense is provided by optical or capacitive pickup. The source of power is a laser stimulated photo-diode. This means that optical access to the silicon layers is required up to the time of measurement. Within these ground rules, a large variety of active structures can be designed to obtain information related directly to device parametric, process, and product variability. In the normal context of yield management this data can be interpreted in combination with that from other sources to provide very early feedback with no scrap or special run cost.

Ring oscillators (ROs) specifically designed for process centering and monitoring of variability have provided a direct link to contact measured parameters on product wafers[5,6]. It is especially advantageous that the same product-like RO design can now be placed in-die with functional and electrical independence of any product structures. Design of a RO to give equivalent results to a

traditional design as used in scribe or product involves primarily the removal of control elements and frequency converters, adjustment for period and power, and addition of non-contact power/sense devices. Restyling a product cell layout can usually facilitate measurement at M1 (first metal) without disturbing any of the device level structures that determine performance tracking with the equivalent probed or product structures while maintaining the density and other requirements for processing subsequent layers.

The ”standard” RO set included in most scribes (INV, NAND, NOR in various configurations generally related to design models) can be extended with designs sensitive to particular parametric or process variations[4,5,6]. The derived data can be device related: effective channel length variation, local random fluctuation, drive strength, threshold, for example. They can also be designed and placed to monitor lot, wafer, and die variations of process specific information: gate/channel interactions, stress effects, rapid anneal, etch control, pattern density, and other exposure effects (including effectiveness of optical proximity correction and resolution enhancement modifications), interconnect performance characteristics and so on).

With somewhat more demand on in-die placement, ROs can be fashioned from arbitrary segments of product logic that mimic critical paths in both function and physical design. These can use any level of interconnect needed to obtain pertinent and timely process or product performance data, as long as the power/sense structures have optical clearance to the surface of the wafer at the time they are measured. It is even possible to repeat measurements of any targets at subsequent process steps if this requirement is met.

In addition to using standard masks and process steps and providing for optical clearance at the time of measurement, the main design requirements for the test targets are: designing for appropriate operating frequencies, assuring adequate power and voltage stability, and verifying the behavior of the designs through model and hardware validation. With regard to frequency, RO operation and signal sense are done at product-like speed. Recalling that the operating frequency of a RO can be estimated as 1/ [2NT], where N is the number of delay blocks and T is the average delay per block, the active area of the test targets can be reduced by decreasing the number of blocks. There are other considerations for reducing the size of the RO, but presently an upper limit of 2GHz is easily accommodated and is not a fundamental limit of the PBM technique. A further note on reducing size: it is important to include evaluation of random local variability in the validation process. Sensitivity to random (versus systematic) variation can be an explicit design goal or a confounding statistical factor to be attenuated, both are possible with PBM.

The active test circuitry is powered by a photodiode illuminated by a focused laser beam. The diode layers aer formed using combinations of implants defined for a given technology. These typically generate between 700 to 800 mV of open-circuit voltage, and approximately 1000 µA of short-circuit current. The voltage generated by the photodiodes is primarily a function of the implant concentrations, junction temperature, and junction current density, which is a function of load, diode area, and laser illumination intensity. The implant profiles in modern CMOS processes are extremely well controlled (consider the precision required to produce the transistors being formed at the same time). Controlling the laser focus and power yields a design space where matching current density to load requirements sets a very stable and repeatable operating point. The reduction in CMOS operating voltages and power requirements with advancing technology nodes allow significant data can be obtained at these voltage and current levels.

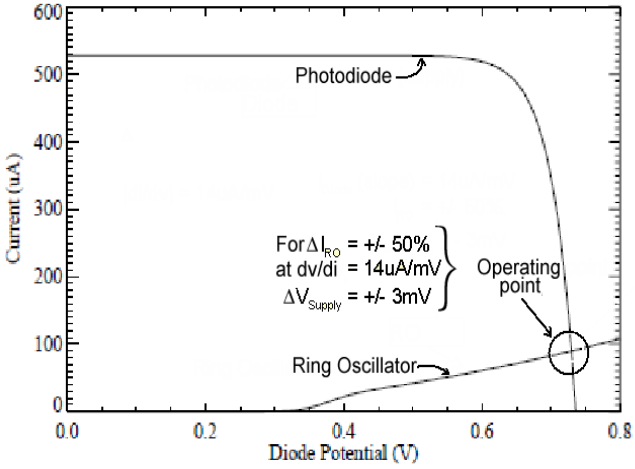



Fig. 3 Operating point and load line

Fig. 3 shows an example of a typical illuminated photodiode I-V (terminal current vs. voltage) plot with a RO load line o (in this case, current out of the anode and into the RO is positive). At the operating point (where the two curves intersect), the RO is drawing about 90uA at 730m.\V. The slope at that point is -14uA/mV resulting in a supply voltage varation of about ±3mv for a 50% variation in RO current draw. By operating circuits in the steep exponential portion of the diode I-V curve, as in this example, ,the diode alone provides good voltage regulation. Standard deviations of around 3mV in open circuit voltage are typical when probing photodiodes across an entire wafer. Which leads to the final part of implementing PBM: the photodiode design options and some variability characteristics of the devices to be monitored need to be evaluated for the specific technology to which PBM is being applied.

The validation process is similar to verifying any standard device or modification thereof (ESD, bipolar, resistor, fuse, etc.). Reasonable combinations of standard diffusions and mask layers are chosen from the standard process definitions for each technology whereby design rules can be derived from initial characterization results. The hardware verification is accomplished by adding a small set of characterization devices with contact pads into unused scribe areas as normal production mask updates are made and thereby introducing them into normal manufacturing work flow with full exposure to the process they will be used to monitor. Validation of measured performance is provided by the same structures. The small size and independence of the functional elements mean that minimal resources are required for this introduction.

### Experimental Results

The immediate question for the user is to ascertain how closely correlated and representative are the PBM results compared with the contact measurements as obtained at final test (end-of-line) or occasional scribe measurements of the same structures and circuit libraries.

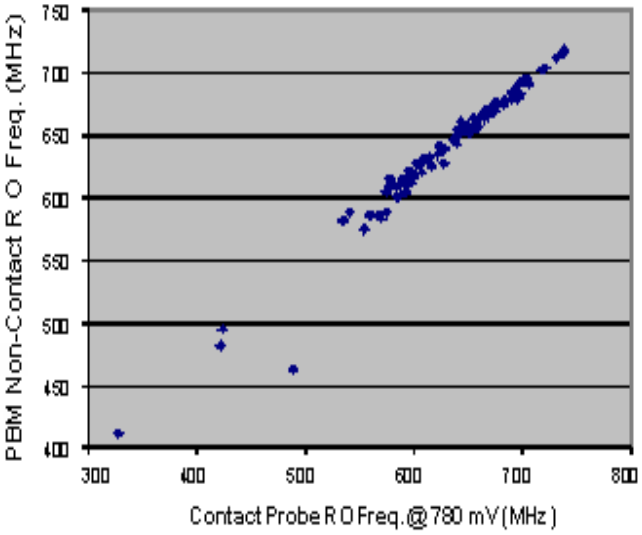



Fig. 4 Non-contact vs. Probed RO measurement (45nm SOI)

The comparison of results from probed structures versus those obtained from the same structure using non-contact power is demonstrated in a scatter plot, Fig 4 (above). These results, taken from a 45nm SOI product wafer, demonstrate the high correlation between the mechanically probed measurements and those from PBM. Fig. 5 shows the same results at the wafer level, and includes the key statistical data, comparing the mean and distribution of the contact vs. non-contact (PBM) measurements. This also demonstrates the excellent spatial correspondence. (The intersection of the red and green areas is the wafer mean with the green region representing faster devices.)

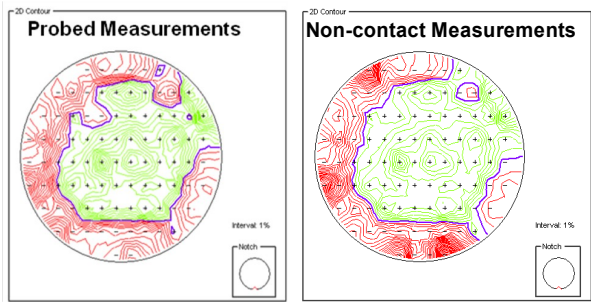



Fig. 5 Wafer map of contact probe vs. PBM non-contact results

Fig. 6 shows the close correlation between electrical Critical Dimension (eCD) measurements of polysilicon gate stripe and those using the same samples (in RO test-structures) measured with PBM on two separate wafers from the same wafer run. PBM identified the slower (larger eCD) wafer from the "faster" wafer (smaller eCD). For each wafer a close correlation between each site and associated CD was also realized.

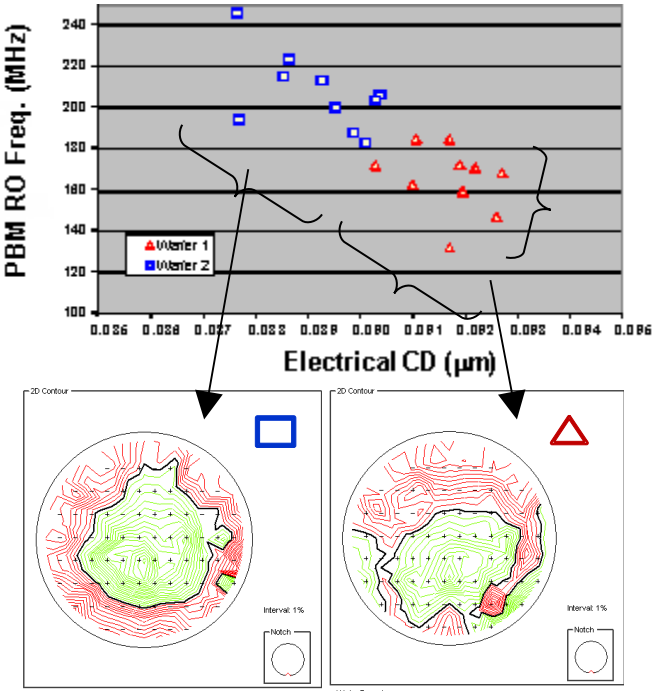



Fig. 6 PBM and eCD correlation (90nm Bulk)

These results show that PBM could enhance or replace eCD with measurements on actual product-like structures (versus the resistivity measurements on polysilicon lines typically used for eCD) and, in turn not only ascertain and control the wafer's patterning processes, but also provide a direct

measure of the effective channel length (Leff) variation – which in turn provides information on the actual physical and electrical construct and constituents of the gate.

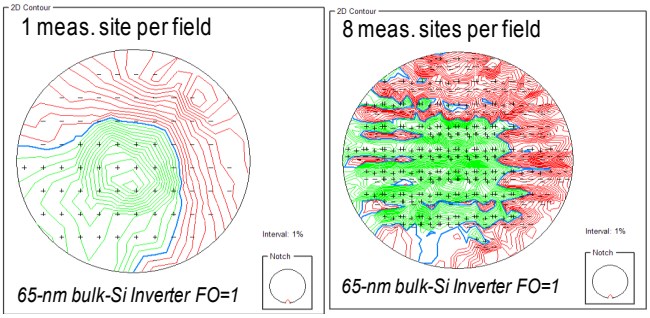



Fig. 7 Single-Point (scribe) vs. Multi-Point PBM

Fig. 7 demonstrates the comparison between single site (per-die) measurements versus multi-site (in-die) measurement on a multi-product wafer (MPW). The single measurements could be viewed as scribe-like results, whereas the multi-site wafers indicate the in-die variability that is ultimately and directly related to performance yield (binning). The average (mean) is the same between the two measurements, whereas the distribution of the results shows 75% more (and measurable) variation is present within the die. These results give comparable resolution, on product, as those usually obtained from test chips or other extensive test element groups [7,8] and, by virtue of the ability to closely integrate the PBM targets with the product circuits, prediction of performance can be made per chip and per wafer.

## Conclusion

A pad-less, non-contact CMOS circuit performance-based metrology system and technology that enables the monitoring and control of process induced variability across wafer and within the product die has been described. The measurement system includes the design and implementation of non-contact power activation and signal detection structures for use in both scribe and within the product die on production wafers. This enables continuous statistical characterization of product performance variation. The use of product-representative and process-sensitive test-structures enhances the value of physical and parametric measurements presently obtained episodically or for tool control or at final wafer test. The measurements can be made as early as first interconnect without wafer scrap, special cleaning, or special masks, and can be expanded or repeated at later steps. Experimental results comparing non-contact and probe-contacted measurements show very good agreement for both bulk-Si and SOI scaled technologies.

These results demonstrate that the non-contact performance-based measurement technique is robust, scales favorably with decreasing device dimensions, and can be utilized in advanced manufacturing lines to directly measure in-die, cross wafer, and cross-line variability. By adopting this technique, critical early process learning can be carried into volume production giving early visibility to bin-yield from first connectivity onward. These features enable accelerated yield ramp, maintain high productivity, and reduce costs.

## References

[1] Duane S. Boning, Karthik Balakrishnan, Hong Cai, Nigel Drego, Ali Farahanchi, Karen Gettings, Daihyun Lim, Ajay Somani, Hayden Taylor, Daniel Truque, Xiaolin Xie, "Variation", *Proceedings of the 8th International Symposium on Quality Electronic Design (ISQED) 2007,* pp. 15-20.

[2] Sharad Saxena, Christopher Hess, Hossein Karbasi, Angelo Rossoni, Stefano Tonello, Patrick McNamara, Silvia Lucherini, Seán Minehane, Christoph Dolainsky, and Michele Quarantelli, "Variation in Transistor Performance and Leakage in Nanometer-Scale Technologies", IEEE *Transazctions on Electron Devices*, vol. 55, no. 1, pp. 131-143, Jan. 2008.

[3] A. H. Gabor, T. Brunner, S. Bukofsky, S. Butt, F. Clougherty, S. Deshpande, T. Faure, O. Gluschenkov, K. Greene, J. Johnson, N. Le, P. Lindo, A. P. Mahorowala, H.-J. Nam, D. Onsongo, D. Poindexter, J. Rankin, N. Rohrer, S. Stiffler, A. Thomas and H. Utomo, "Improving the Power-Performance of Multicore Processors Through Optimization of Lithography and Thermal Processing", *Proc. of SPIE,* vol. 6521, pp. 65210K, 2007.

[4] Mark B. Ketchen, Manjul Bhushan, "Product-representative 'at speed' test structures for CMOS characterization", *IBM Journal of Research and Development*, vol. 50, no. 4/5, pp. 451-468, 2006.

[5] Kelin J. Kuhn, Chris Kenyon, Avner Kornfeld, Mark Liu, Atul Maheshwari, Wei-kai Shih, Sam Sivakumar, Greg Taylor, Peter VanDerVoorn, and Keith Zawadzki, "Managing Porcess Varaition in Intel's 45nm CMOS Technology", *Intel technology Journal*, vol.12, no.2, pp. 93-109, 2008.

[6] Bhushan, M. Gattiker, A. Ketchen, M.B. Das, K.K. , "Ring oscillators for CMOS process tuning and variability control", *IEEE Transactions on Semiconductor Manufacturing,* vol. 19, no. 1, pp. 10-18, Feb. 2006.

[7] Hiroo Masuda, Shin-ichi Ohkawa, Atsushi Kurokawa, Masakazu Aoki, "Challenge: Variability Characterization and Modeling for 65- to 90-nm Processes", *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 593-599, 18-21 Sept. 2005.

[8] M. Karthikeyan, A. Gasasira, S. Fox, G. Yeric, M. Hall, J. Garcia, et al., "Development and Use of Small Addressable Arrays for Process Window Monitoring in 65nm Technology," *Proc. IEEE Conf. Microelectronic Test Structures (ICMTS)*, Mar. 2007.

# Test Structure for High-Voltage LD-MOSFET Mismatch Characterization in 0.35 um HV-CMOS Technology

Werner Posch, Christian Murhammer, Ehrenfried Seebacher

*austriamicrosystems* AG
A-8141 Unterpremstätten, Austria
werner.posch@austriamicrosystems.com

## ABSTRACT

A characterization setup for high voltage (HV) LD-MOSFET mismatch and variability determination is presented. The according test chip was successfully realized in 0.35 um HV-CMOS technology. Devices are aligned in rows and columns for gate and drain bias multiplexing and special HV-switches for voltages up to 50V are controlled by externally generated digital signals. Automatic DC measurements can be performed on up to 4992 HV-NMOSFETs, providing variability data for both, short and long distance matching characterization.

## INTRODUCTION

The wish to handle high voltages in e.g. in automotive electronics, display drivers, switched power supplies, etc. on the one hand and to increase integration density on the other hand led to development of high voltage devices on chip. Especially the HV lateral diffused FET became quite popular, because it could be implemented in standard CMOS processes using only a small number of additional masks. The high voltage capability induced deviations of drain current behavior compared to a standard FET device. Especially an increased on-resistance and a kind of quasi-saturation at high gate-source voltages was typically found for short channel HV LD-FETs, as demonstrated in the output characteristics in Fig. 1. Additionally to the standard needs for simulation (DC, AC), mismatch considerations and simulations are important parts during the design phase and help to predict and optimize yield. Determining both, short distance and far distance mismatch parameters for application in Monte Carlo SPICE simulation was the motivation to design this new test chip.

*HV-LDMOS device behavior*

The device cross-section is shown in Fig 2. This introduction presents a short overview of primary high-voltage extensions applied to a standard FET structure, more information can be found e.g. in [1]-[4], where the device physics of high voltage effects is described in detail, especially focusing on the reasons for the quasi saturation behavior.

A) First, an increased junction-breakdown voltage of the drain diffusion is achieved if using a deep drain well. The well doping is of course much smaller than the contact diffusion doping and the radiuses of the cylindrical and spherical junctions at the borders are large.

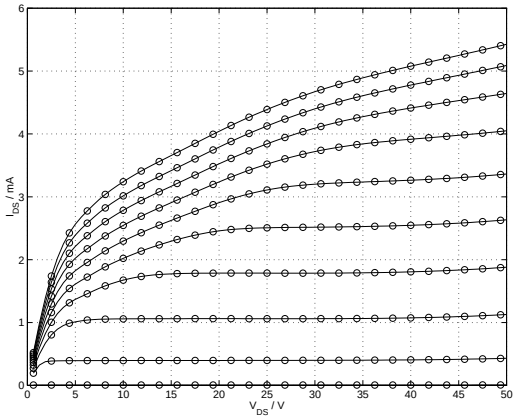



Fig. 1. Output characteristic of an isolated HV-NMOS, W/L=10/0.5, drain current $I_{DS}$ vs. drain source voltage $V_{DS}$ for $V_{GS}$ = 2.9 to 20V, step 1.9V.

Both, doping considerations and geometrical considerations increase the breakdown. As small on-resistance and high breakdown voltage are contrary effects, the actual drain well doping is adjusted as large as possible to provide the smallest on-resistance achievable for a given breakdown voltage.

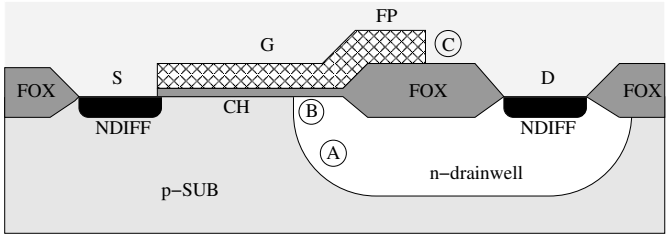



Fig. 2. Substrate placed HV-NMOS: S..source, G..gate, D..drain, FOX..field oxide, NDIFF..contact diffusion, CH..channel, FP..field plate

B) Second, the gate length is extended beyond the body-drainwell junction, which increases the junction breakdown voltage. The gate is a kind of field plate and bends the electric field in a way that the critical field strength occurs at increased drain source voltages, commonly known as RESURF effect [4].

C) Third, as there are high voltages at the drain, the electric field at the end corner of the gate electrode becomes quite large due to the small radius. Therefore field oxide or shallow trench isolation is used in order to separate the critical gate region and the drain region.

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

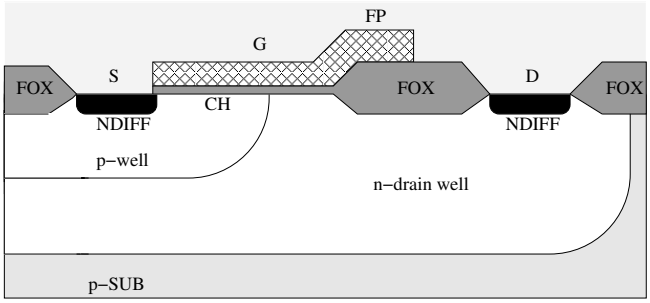



Fig. 3. Isolated HV-NMOS: S..source, G..gate, D..drain, FOX..field oxide, NDIFF..contact diffusion, CH..channel, FP..field plate

Figure 2 and Fig. 3 show different realizations of HV-NFETs. The structure of the isolated type is more complex than the substrate placed device but allows free body well biasing. This is greatly appreciated in high-side switch applications, where body and source potentials may be far beyond the substrate ground potential. In most cases source and body terminals are shorted in order to prevent threshold voltage increase and to receive minimum on resistance, respectively. As substrate is the body terminal of the substrate placed type, applications are limited to the allowed operation conditions of the source-substrate junction.

*Matching considerations for HV-FETs*

Mismatch characterization considers stochastic, time independent variations of devices having similar layout [5], [6]. A lot of analog circuits presume perfectly matched devices and performance may drastically be reduced if mismatch effects are ignored. Accurate mismatch modeling helps to estimate yield during the design phase and reduces costs due to overestimated design margins. Generally, for mismatch modeling an enormous number of measurement data has to be acquired to apply statistical analysis for parameter extraction. Test structures usually contain several matched pairs having different device size. They can share common electrodes and Kelvin-probe measurements compensate contact wire resistances. Measurements are performed on large arrays using servo controlled micro probers or automated switching test equipment. Multiplexing of device terminals reduces the number of pads efficiently. Even for HV-FETs, where the device size is about half a probe-pad size, area is reduced by an approximate factor of four if using switching on chip. Measurements can be performed automatically and the device under test is addressed by digital signal sequence. Analog switches are transmission gates, where the electrodes of the parameter analyzer are connected to a device under test. Contrary to earlier realized structures for standard FETs [7]-[11], the test chip introduced in this paper must be able to handle voltages up to 50 V. Special transmission gates have been designed for gate and drain voltage multiplexing. Each transmission gate is controlled on chip by the output of a shift register. The external parts of the measurement setup are shown in Fig. 4.

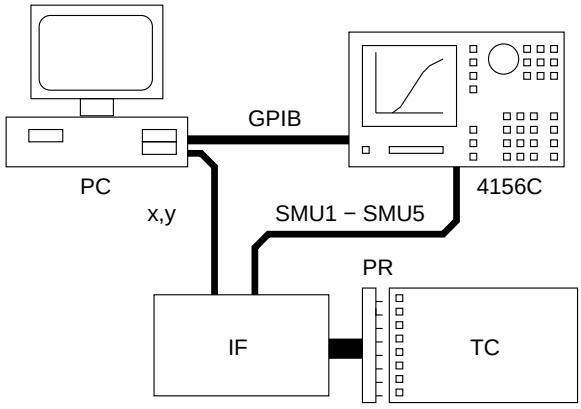



Fig. 4. Measurement Setup: GPIB…instrument interface, 4156C…parameter analyzer, xy… device selection, IF…interface board, PR…probe card, TC…test chip

## TEST STRUCTURES

Two test structures were designed and characterized in detail. They have identical pad assignment and can be controlled with the same external measurement hardware. The first circuit is a one dimensional prototype for evaluation of high-voltage transmission gates for gate bias multiplexing. In the second chip the concept had been extended by adding additional switches for drain terminal multiplexing. An array of 24 times 208 high voltage NMOSFETs is available for characterization.

*Test Structure for Gate Bias Multiplexing*

In order to evaluate the overall concept, a fully functional but small test structure has been realized. Fig. 5 shows the circuit schematics. The eight bit shift register (VDD=3.3V) is controlled by external signals CLK, DO, RES from the interface board. Each output forces a high voltage switch to connect the gates of the devices either to the SMU of the parameter analyzer or to ground. Source, Drain and Bulk terminals are directly connected to the according SMU of the parameter analyzer Agilent 4156C.

With this structure 2 x 8 HV-FETs can be measured by using two separate drains D1 and D2. To compare measurement results depending on x and y directions, a pair of identical transistors is arranged in x and y direction, resulting in four identical transistors close to each other. The device sizes are W/L=40/10 and 40/$L_{MIN}$ for both, the isolated transistor type and the substrate placed transistor type. To prevent large voltage drops on the metal wiring, transistors with large current driving capability are positioned close to force pads. Additionally, stacked metal is used to reduce the sheet resistance of the metal interconnect. The layout is made fully symmetrical to the source metal line, leading to equal distances to the sense points and to reduced systematic offsets due to contacting.

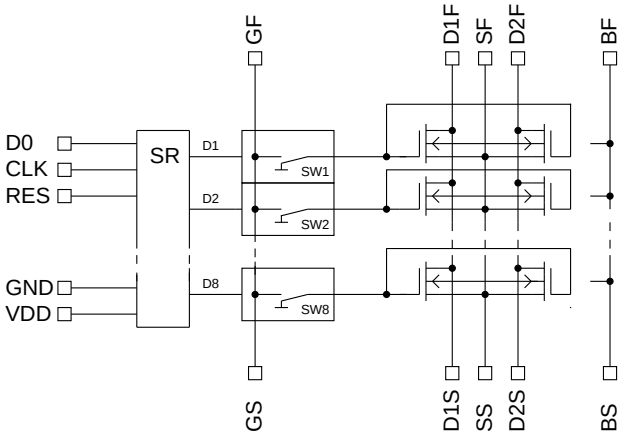



Fig. 5. Simplified schematic circuit of the first test chip. SR…shift register, SW…HV-switch, D…digital output of shift register, top: force contacts, bottom: sense contacts.

A microphotograph of the chip is shown in Fig. 7. Both types of investigated transistors have 48 nm gate oxide with maximum gate-source voltage $V_{GSMAX}$ = 20V operating condition, which is the minimum requirement for the switches. Even because of some high voltage design considerations, the switch functioning has been successfully realized within a range of 0 to 50V. As can be seen in the schematics, no additional supply voltage is needed for a switch. Instead, the gate SMU has been used. This gate SMU current can be monitored during the measurement too, giving additional information about the correct functioning of the switch.

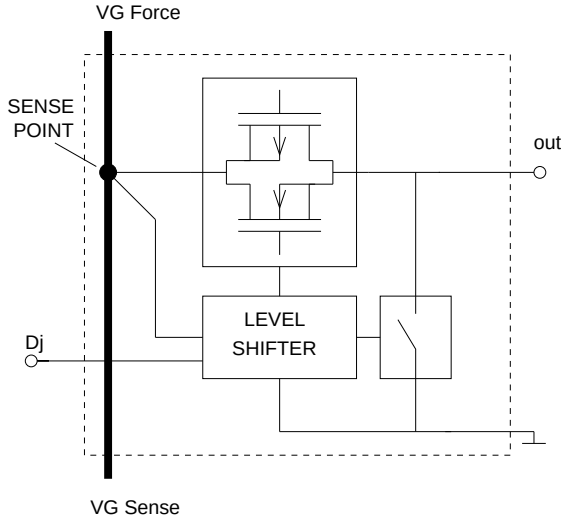



Fig. 6. Simplified schematic circuit of the gate switch

The gate switch [12] has to provide only small load currents for gate capacitances. Transient SPICE simulation verified that W/L=10um/$L_{MIN}$ device sizes for n-HV-FET and p-HV-FET transmission gate transistors are a proper choice. In order to minimize switch variability the connections to transmission gate transistors and the supply of the level shifter circuit have their origin at a well defined sense point, as illustrated in the block schematics of the switch in Fig. 6.

As mentioned, this structure was mainly used for evaluation of the concept, the switches and the other components. It was additionally intended for ramping up the measurement setup and software. This small circuit can be easily extended by adding shift registers and switches, which is shown in the next section.

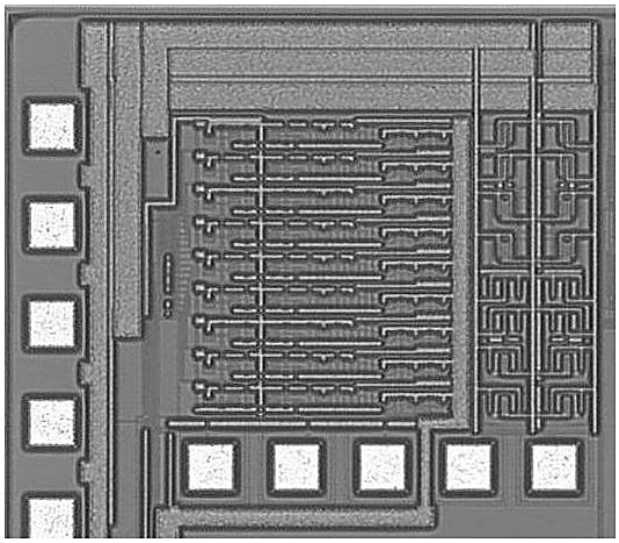


Fig. 7. Microphotograph of the first test chip. right: devices under test, center: HV-switches and shift register, bottom: contact pads for switch characterization.

*Test Structure for Gate and Drain Bias Multiplexing*

The formerly described switching unit for gate voltages has been arranged 26 times in series, giving a total number of 208 equidistant lines for gate bias multiplexing. Using matrix structure, a column represents a common drain electrode. One common SMU for drain biasing is shared to 24 columns by using drain switches giving a total number of 4992 devices accessible on the chip. Each column consists of devices of identical size. The schematic circuit is shown in Fig. 8. Four different types of transistors are chosen for investigation: two isolated transistors with $V_{DSMAX}$ = 20V and $V_{DSMAX}$ = 50V and two substrate placed NMOS having $V_{DSMAX}$ = 50V. Six different device sizes are realized for each type to provide size dependent variability data for modeling [5]. The maximum drain voltage is $V_{DSMAX}$ = 50V. Therefore the operating range of the drain switch must be between 0 and 50V. Additionally one has to take into account the maximum drain currents of the devices under test for a proper design of the transmission gate. In contrast to the gate switch, the drain switch cannot be supplied via the drain SMU. The additional voltage is provided by the high power SMU. The monitoring this current gives information about the correct functioning of the drain switch.

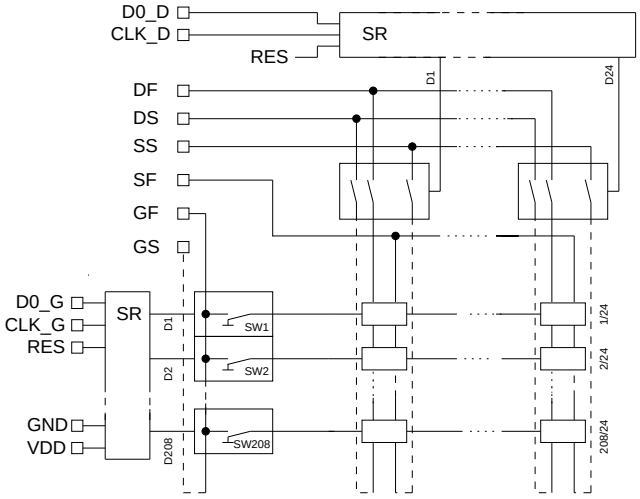



Fig. 8. Simplified schematic circuit of the second test chip. SR…shift register, SW…HV-switch, D…digital output of shift register, D0_D, CLK_D, D0_G, CLK_G digital signals for selecting drain and gate switch. Terminals: second letter stands for (S)ense or (F)orce.

The transmission gate is designed using a single isolated n-channel HV-FET, which operates as a high-side switch. Generally, it is possible to adapt the concept of the gate switch for driving high currents if device widths of switching transistors are increased and an independent supply voltage is provided externally. The disadvantage of this variant is that the n-HV-FET is switched on for bias voltages up to about VDD. For higher drain bias, all the current must pass through the p-HV-FET, which has a smaller current drive capability compared to an n-HV-FET of equal width. Therefore a high-side n-HV-FET switch circuit was designed where no p-channel transistor is needed. The large current drive capability of the n-FET reduces area consumption drastically.

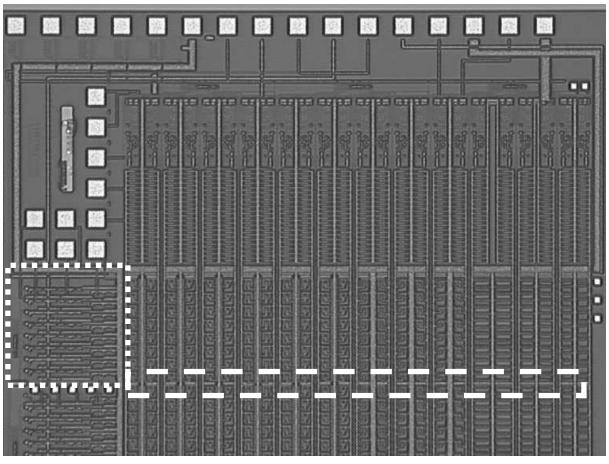


Fig. 9. Top part microphotograph of the second test chip. Top: contact pads, bottom: devices under test, left: gate switches and contact pads for characterization of the drain switches. dotted square: compound of eight gate-switches, dashed rectangle: horizontal, stacked ground metal wiring. Chip size: ~2 x ~10 mm

## RESULTS

Generally, transfer characteristics ($I_{DS}$-$V_{DS}$) of a matched pair are measured for mismatch determination and direct or indirect extraction methods are used for mismatch parameter extraction. Direct methods extract device parameters e.g. $V_{TH}$, $K_P$, etc. for each device and statistics provides results for variations and mean values. Indirect methods use a proper variance model and nonlinear optimization to directly extract variances out of $\sigma(\Delta I_{DS}/I_{DS}) = \sigma( 2(I_{DS1}-I_{DS2}) / (I_{DS1}+I_{DS2}) )$ versus $V_{GS}$ characteristics.

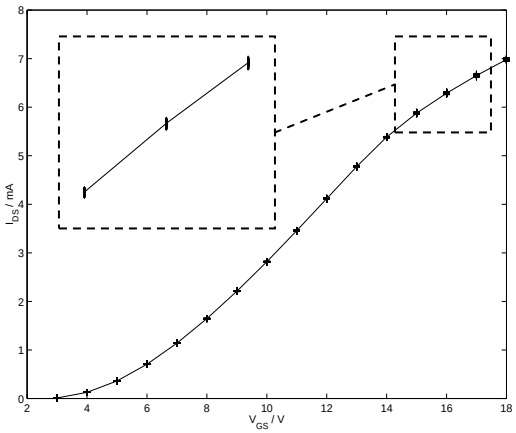



Fig. 10. Transfer characteristic data of 208 isolated transistors. Drain current $I_{DS}$ vs. gate-source voltage $V_{GS}$. +...measurement data, -…mean values. It shows increased mismatch in quasi-saturation region. $V_{GS}$ = 3 to 18V, 16pts., $V_{DS}$=20V, W/L=30/3.

A transfer characteristic data set of 208 isolated transistors ($V_{DSMAX}$=50V) is shown in Fig. 10. These measurements verified the effect of increased drain current mismatch of isolated transistors found for high $V_{GS}$ in the quasi-saturation regime. Fig. 11 shows this behavior clearly. It contains $\sigma(\Delta I_{DS}/I_{DS})$ data of transistors close to each other (even an odd rows) for linear and saturation operation regimes with zero and nonzero bulk bias applied. $V_{DS}$=0.1V, 20V; $V_{BS}$=0V and $V_{BS}$ = -0.5V.

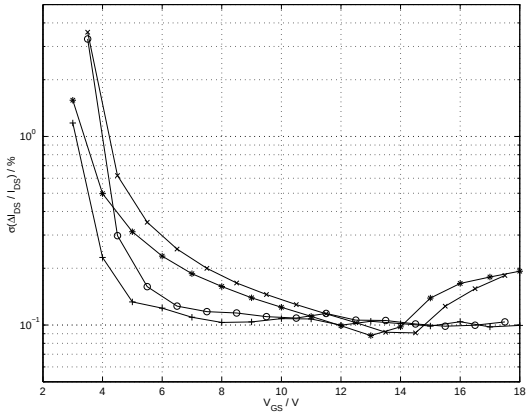



Fig. 11. Drain current mismatch $\sigma(\Delta I_{DS}/I_{DS})$ vs. gate-source voltage $V_{GS}$ for (3-4) + (5-6) pairs. *…$V_{BS}$=0V, $V_{DS}$=20V, +…$V_{BS}$=0V, $V_{DS}$=0.1V, o…$V_{BS}$= -0.5V, $V_{DS}$=0.1V, x…$V_{BS}$= -0.5V, $V_{DS}$=20V.

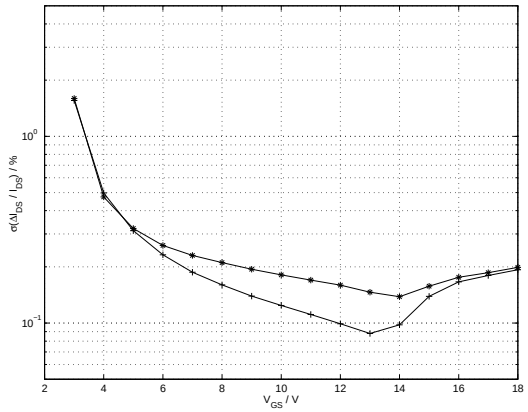



Fig. 12. Drain current mismatch $\sigma(\Delta I_{DS}/I_{DS})$ vs. gate-source voltage $V_{GS}$ for +: (3-4)+(5-6) pairs and *: (1-2)+(7-8) pairs.

Further the matching of devices controlled by a compound of eight switches, shown in the dotted square in Fig. 9, was investigated. As a result, different mismatch behavior was found for devices inside and for devices on the border of a block of eight devices. Figure 12 illustrates mismatch of (3-4) together with (5-6) pairs and larger mismatch of (1-2) + (7-8) pairs. This tendency was reproduced for data of fife different sites. A reason could be the influence of backend wiring [6], especially the horizontal stacked metal ground line, which is the only difference between the two types of pairs. The mismatch data of (3-4)+(5-6) pairs correspond to results measured on standard matched pair structures. In order to investigate distance dependent threshold voltage variation, differences $\Delta V_{TH} = V_{TH1}-V_{TH2}$ of threshold voltages of two transistors close to each other and pairs separated by a distance d = 104 * 43.6 um = ~4.5 mm, are shown in the histograms in Fig. 13 and Fig. 14. The standard deviations σ and mean values μ increase from $\sigma(\Delta V_{TH})_{pair}$ = 3.5 mV, $\mu(\Delta V_{TH})_{pair}$ = 0.02 mV to $\sigma(\Delta V_{TH})_{dist}$ = 4.0 mV and $\mu(\Delta V_{TH})_{dist}$ = 1.4 mV. The shift of the mean value indicates a $V_{TH}$ gradient.

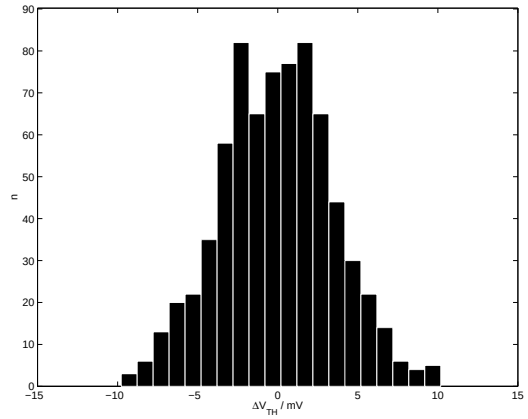



Fig. 13. Histogram of isolated transistor short distance threshold voltage mismatch $\Delta V_{TH}=V_{TH1}-V_{TH2}$. (odd and even rows, seven sites), W/L=30/3, $V_{BS}$=0V, $V_{DS}$=20V.

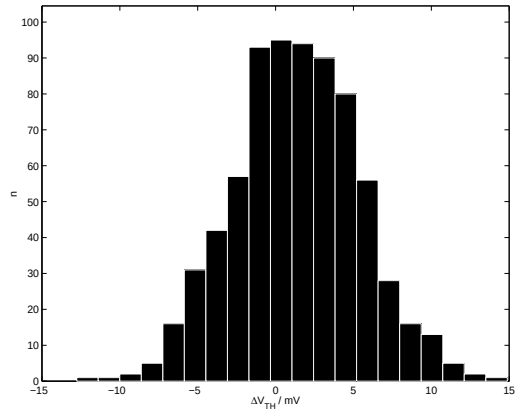



Fig. 14. Histogram of isolated transistor far distance threshold voltage mismatch $\Delta V_{TH}=V_{TH1}-V_{TH2}$. (1: 1 to 104, 2: 105 to 208, seven sites), W/L=30/3, $V_{BS}$=0V, $V_{DS}$=20V.

Results even show, that the influence of distance to σ is small. Matching is mainly determined by short distance pair mismatch parameters. Additionally, an autocorrelation analysis showed that threshold voltages are uncorrelated and randomly distributed.

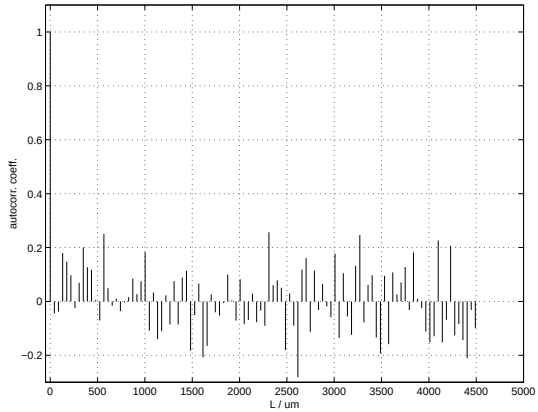



Fig. 15. Distance autocorrelation function of threshold voltage. W/L=30/3, $V_{BS}$=0V, $V_{DS}$=0.1V.

## CONCLUSION

A test setup for determining matching behavior of high voltage n-channel LD-MOSFETs was presented. Digitally controlling and multiplexing of bias voltages up to 50V on chip allowed automated measurements of DC-characteristics on an array of 4992 HV-FETs. Pair mismatch characterization and distance dependent mismatch characterization of transistor parameters were demonstrated for the threshold voltage of an isolated HV-LDMOS transistor type. The standard deviation of the threshold voltage mismatch increases 15% if distance d = ~4.5mm between two transistors is considered. The offset value due to distance is approximately one third of the standard deviation value.

## REFERENCES


[1] Ballan Hussein, Michel Declercq: "High Voltage Devices and Circuits in Standard CMOS Technologies", Kluver Acadeemic Publishers, Dortrecht, Boston, London 1999.

[2] M. Knaipp, G. Roehrer, R. Minixhofer, E. Seebacher, "Investigations on the High Current Behavior of Lateral Diffused High-Voltage Transistors", IEEE Transactions on Electron Devices, Vol. 51, No. 10, Oct. 2004, pp. 1711-1720

[3] M. Knaipp, J. M. Park, V. Vescolli, G. Roehrer, R. Minixhofer, "Investigations on an Isolated Lateral High-Voltage n-channel LDMOS Transistor with a Typical Breakdown of 150V", Proceedings of the 36th European Solid-State Device Research Conference 2006, ESSDRC 2006.

[4] Z. Parpia, C. A. T. Salama, "Optimization of RESURF LDMOS Transistors: An Analytical Approach", IEEE Transactions on Electron Devices, Vol. 37, No. 3, March 1990, pp. 789-796.

[5] M. Pelgrom, C. Duinmaijer, A. Welbers: Matching Properties of MOS Transistors, IEEE Journal of Solid-State Circuits, Vol. 24, No. 5, Oct. 1989, pp. 1433-1440.

[6] H. P. Tuinhout, M. Vertregt, "Characterization of Systematic MOSFET Current Factor Mismatch Caused by Metal CMP Dummy Structures", IEEE Transactions on Semiconductor Manufacturing, Vol. 14, No. 4, Nov. 2001, pp. 302-310.

[7] K. M. G. V. Gettings, D. S. Boning, "Study of CMOS Process Variation by Multiplexing Analog Characteristics", IEEE Transactions on Semiconductor Manufacturing, Vol. 21, No. 4, Nov. 2008, pp. 513-525.

[8] K. Agarwal, J. Hayes, S. Nassif, "Fast Characterization of Threshold Voltage Fluctuation in MOS Devices", IEEE Transactions on Semiconductor Manufacturing, Vol. 21, No. 4, Nov. 2008, pp. 526-533.

[9] H. Klimach, C. Galup-Montoro, M. C. Schneider, "A Setup for Automatic MOSFET Mismatch Characterization under a Wide Bias Range", NSTI-Nanotech 2007, Vol. 3, 2007.

[10] U. Schaper, C. G. Linnenbank, R. Thewes,"Precise Characterization of Long-Distance Mismatch of CMOS Devices", IEEE Transactions on Semiconductor Manufacturing, Vol. 14, No. 4, Nov. 2001, pp. 311-317.

[11] L. Portmann, C. Lallement, F. Krummenacher, "A High Density Integrated Test Matrix of MOS Transistors for Matching Study", Proc. IEEE Int. Conference on Microelectronic Test Structures, Vol. 11, March 1998, pp. 19-24.

[12] C. Murhammer, "Design and Simulation of a Mixed Signal Testchip in a 0.35µm HV-CMOS Technology for High Precision DC Measurements", University of Applied Sciences, Sankt Poelten, 2006.


## AUTHOR'S BIOGRAPHIES

Werner Posch received the M.S. degree in physics from Graz-University of Technology in 2002. Currently, he is working as a process characterization engineer at austriamicrosystems AG and is responsible for high voltage LD-MOS transistor SPICE modeling and mismatch characterization.

Christian Murhammer received the M.S. degree in scientific computing from University of Applied Sciences, Sankt Poelten, in 2006. He gained research experience in HV-CMOS circuit design during his final year project at austriamicrosystems AG. Since 2006, he is working as a design engineer at Schiebel Elektronische Geraete GmbH.

Ehrenfried Seebacher received the M.S. degree in physics from Graz-University of Technology in 1993. Since 1999, he has been the section manager of austriamicrosystems' process and device characterization group, which is responsible for COMPACT modeling, process characterization, verification run-sets, and SPICE simulator support.

# Automated Test Structure Generation for Characterizing Plasma Induced Damage in MOSFET Devices

Thomas Zwingman, Ann Gabrys, and Andrew J. West

National Semiconductor Corporation
2900 Semiconductor Drive, Santa Clara, CA, USA, 95052
thomas.zwingman@nsc.com

## ABSTRACT

Test structures used to study the effects of plasma induced damage are complex and time intensive to design; performance problems due to poorly designed components of the structure often confound the desired result. This paper presents a parameterized and hierarchical antenna test structure template that enables the user to characterize the test structure performance and identify safe design guidelines early in process development. The template is implemented in a system that automates structure generation, placement, routing, and test plan development.

## INTRODUCTION

Plasma induced charging is a significant mechanism that causes damage to the gate oxide of Metal Oxide Semiconductor (MOS) devices. MOS transistors connected in parallel with antenna structures are used as test devices to assess damage due to plasma exposure from back-end processing [1]. In many cases, the in-process electrical overstress occurs from the buildup of an electrical charge on the gate resulting in current flow through the oxide [2]. A typical antenna protection diode can provide an alternative leakage path for the plasma induced charge [3,4].

Due to their relative complexity, antenna test structures used to characterize the plasma induced leakage are typically modified layouts from previous test chips. This approach limits the number and scope of the experimental parameters and structures that can be analyzed. In this work, a hierarchical template with 18 degrees of freedom was constructed for use within the layout automation system outlined in Figure 1 [5]. This template was used to generate, place, and route 60 antenna test structures in an experiment to examine the relationship between the sub-circuit components of the MOS transistor, antenna protection diode, and associated pad connections.

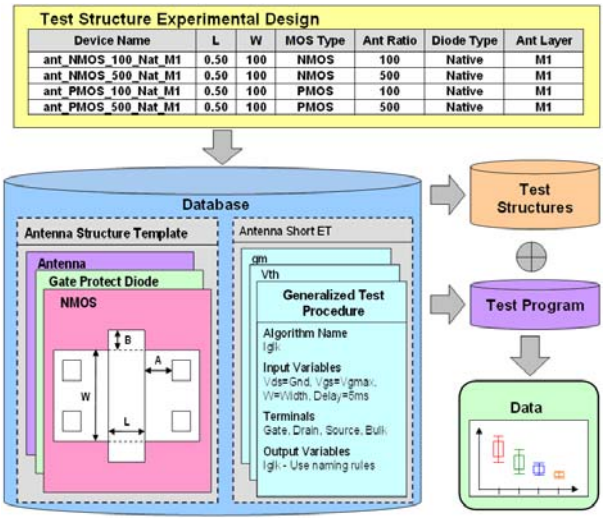



Fig. 1. Automated test structure and test plan development flow.

## AUTOMATED TEST STRUCTURE DEVELOPMENT

The standard antenna test structure consists of a MOSFET gate connected to a large metal antenna. Test chip experiments designed to determine layout design rules investigate different antenna ratios for varying metal/poly layers, gate oxide thicknesses (for multiple gate oxide processes), and antenna protection diode types.

Figure 2 shows two diode protection schemes that can be used to prevent gate leakage from plasma induced damage in back-end processing. The first illustration provides a cross-section of an NMOS transistor with P+/N-Well/P-Sub floating N-Well double-sided diode protection. This has the advantage of allowing positive and negative gate biases to be applied. The second example is shown with the N+/P-Well diode sharing the same P-Well with the NMOS transistor.

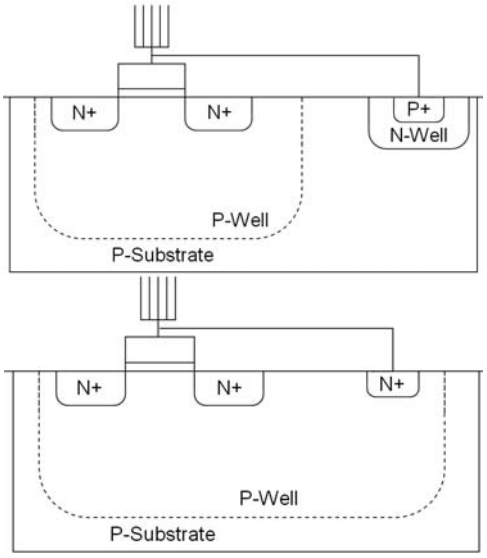



Fig. 2. Variations of protection diode scenarios of inside-well N+/PW and P+/N-Well/P-Sub floating N-Well double sided diodes.

Previously, antenna test structures were constructed from available reference layouts. To implement structure variations, each element in the sub-circuit required individual modifications. As a result, the layout resources and time required increased with the complexity of the experiment and was dependant on the design parameters built into the reference cells.

The template developed in this work combines a MOS transistor, an antenna structure, and a gate protection diode with design parameters available at both the individual component and sub-circuit levels. The complete list of the 18 available parameters is shown in Table 1.

As shown in Figure 1, the user sets all the parameter values defined for the experiment in a single point of entry. Since the antenna template is defined in a hierarchical format, it references existing device templates, such as the MOS transistors and diodes, and passes the defined design parameters accordingly. By accessing a central database, the template references layout design rules to create the elements within the

Table 1. List of the 18 available parameters for characterizing the antenna sub-circuit.

| Device | Parameters (units) | Range of Values |
|---|---|---|
| **MOS Transistor** | MOS Type | NMOS, PMOS |
| | Gate Width (µm) | 0.05-100 |
| | Gate Length (µm) | 0.05-100 |
| **Antenna** | Antenna Ratio | 0-10000 |
| | Antenna Layer | Con, Poly, M1-Top Metal, Via 1 - Top Via |
| | Antenna Type (Area/Edge) | Area, Edge |
| | Finger Contact On/Off | On, Off |
| | Contact Ratio | 0-2000 |
| | Finger Separation (µm) | 0-100 |
| | Finger Width (µm) | 0-100 |
| | Metal Jumper Layer | M1-Top Metal |
| **Protection Diode** | Diode Area ($\mu m^2$) | 0-1000 |
| | Diode Type | Back2Back, Native |
| | Protection Diode On/Off | On, Off |
| | Diode Offset (µm) | 0-30 |
| | Diode Placement | In-Well, Outside Well |
| **Top Level** | Pad Protection Diode On/Off | On, Off |
| | Nominal Bias (V) | 0-50 |

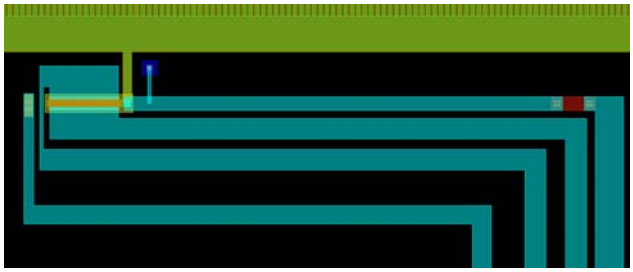


Fig. 3. One example layout of 60 automatically generated for a test chip experiment spanning 18 design variables.

template. Once the test structure definition is complete for all structures defined in the experiment, a high-level algorithm is applied to place and route all structures within test pad modules.

Figure 3 contains the layout for one of 60 unique test structures generated using the parameterized template for a series of experiments spanning all 18 design parameters. This example contains an NMOS transistor connected to a large edge-intensive Metal1 antenna and a separate-well diode. To avoid the plasma induced from the test pad creation, the metal jumper was connected at Metal2.

As is typical for any automation effort, the upfront time and costs must outweigh the overall time to develop the automation capability for the specific application. For this experiment, approximately two weeks of effort was initially invested in the template design and the routing algorithm development. This one-time effort was in part due to the complexity of the intended experiment, but was independent of the size of the experiment. Once the template was created, experiment definition and test chip module generation required one day of effort, again independent of the size of the experiment, and would apply to all successive test chips.

Another key time-saving advantage is the ability to port the template to a new technology. Each device contained within the antenna template hierarchy is constructed referencing layout rules defined in a process-independent manner. As a result, transferring the template to a new process can be completed in a couple of days once the new process layout rules have been entered in the database. Alternatively, modifying the structure with manual layout software between technologies requires significant modification of process layers, layer spacings, device instance references, etc.

To fully realize the time and cost savings, it is necessary to compare the setup time to performing manual layout and routing. Manual layout depends on

the number of sub-circuit components and scales with the size of the experiment. This is an inflexible approach where each structure requires manual updates. The experiment analyzed in this work would have taken two to three weeks of manual layout generation. This effort is repeated for each subsequent test chip and technology unless an exact copy is applicable. In addition, every modification introduces a chance for error. Conversely thru automation, errors can only be introduced during experiment definition and entry.

## RESULTS

Using the automation software, multiple MOS devices were created to determine the optimal device configuration to study the plasma induced damage effects. The technology used in this study was a 90nm dual-well, dual-gate oxide CMOS process with 20Å and 60Å gate oxide thicknesses. The back-end processing consists of six metal layers. The results described represent a subset of the larger experiment to study the effects of protection diode placement relative to the MOS, diode types, device performance degradation, and gate oxide thickness.

Gate leakage measurements were taken in order to determine the effect of placement of the protection diode relative to the MOS transistor. Figure 4 highlights the leakage distribution differences between the placement of a diode inside (same-well) versus outside (separate-well) the P-Well of the NMOS transistor. The separate-well N+/PW diode shows a high gate leakage tail, similar to not having a protection diode in the circuit. Diode placement relative to the well produces a significant response.

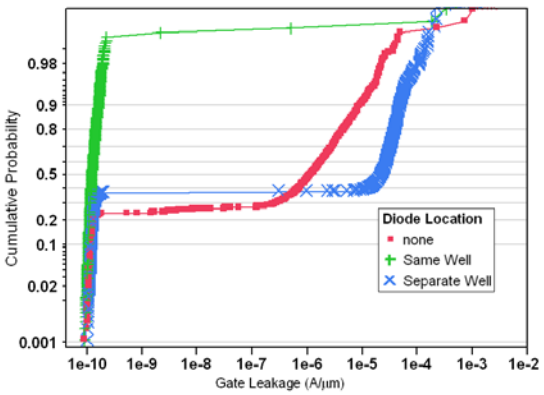



Fig. 4. Gate leakage distributions comparing placement of N+/PW diode in same versus separate-well and with no diode protection. Each structure has a 2000:1 Metal1 to gate area antenna ratio.

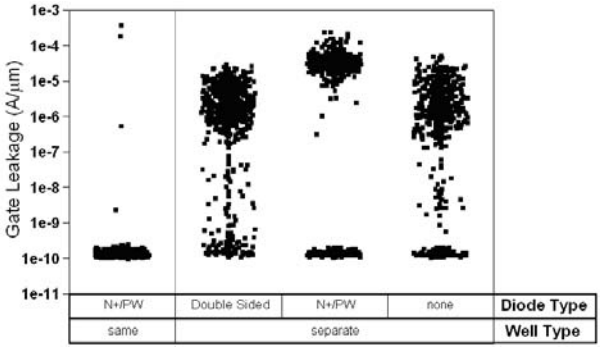



Fig. 5. Gate leakage distribution comparing effects from difference well and diode types. Each structure has a 2000:1 Metal1 to gate area antenna ratio. The diode for the separate-well N+/PW structure is placed in a floating P-Well.

Figure 5 shows the gate leakage difference between a double-sided diode scheme and one with same and separate-well N+/PW diodes for NMOS transistors (a P+/NW diode would be used for a PMOS transistor). A diode placed in a separate-well fails to protect the device regardless of the diode selection. The reduced gate leakage from the same-well N+/PW diode will enable larger antenna ratios for the same size transistor.

Given the amount of plasma induced damage to the MOS gate oxide, the transistor off-state drain leakage ($I_{OFF}$) was correlated to the measured gate leakage. The elevated $I_{OFF}$ data correlates to a high gate leakage value (Figure 6). This implies that the gate oxide was damaged by the plasma induced charge and degraded the performance of the device.

Figure 7 compares the gate leakage probability distributions between fully processed antenna structures of thin-gate and thick-gate oxide transistors, both using a double-sided diode scheme. The gate leakages were normalized given the differences in oxide thickness and show that with identical protection structures, the thick-gate oxide transistors do not exhibit the plasma damage. The diode protection schemes for a 60Å gate oxide thickness cannot be applied to thinner oxides of the order of 20Å.

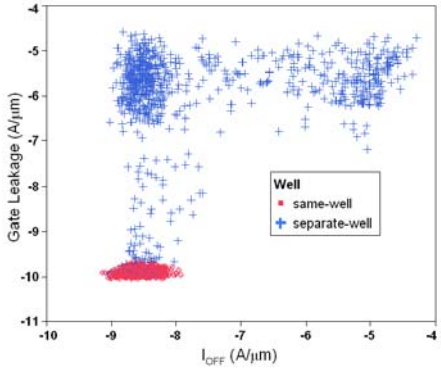



Fig. 6. Scatter plot of off-state drain MOS leakage compared to gate leakage for in-well versus separate-well diode structures.

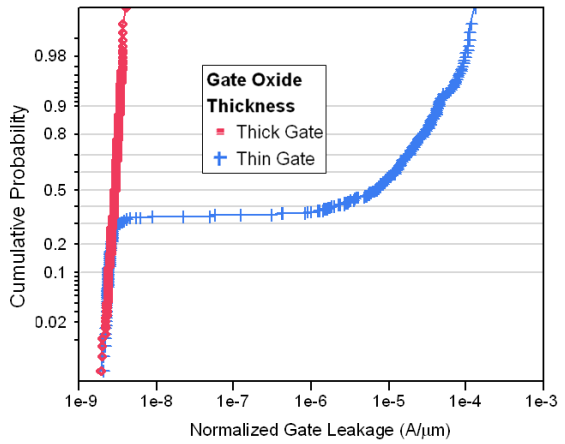



Fig. 7. Normalized gate leakage distributions of thin versus thick-gate oxide MOS antenna structures.

A potential limitation for designing transistors with an N+/PW diode inside the same P-Well for an NMOS is that the diode needs to be placed in close proximity to the gate. Figure 8 shows the gate leakage distributions for a same-well N+/PW diode design varying the distance in the placement of the diode. Distances up to 100µm do not degrade MOS gate leakage performance. This simplifies the experiment for subsequent test chips and helps identify appropriate design rules.

The ease of creating an extensive array of test structures and several experiments utilizing the 18 possible parameter variables allows for a broad analysis of plasma charging effects on a particular process technology. The results can be used to optimize test structure sensitivity, allowing for a reduced structure set for the final process monitor. The advantage of identifying the optimum test structure configuration allows the user to develop an appropriate antenna experiment using data that is not confounded by the MOS device layout. This significant savings of time and test chip area becomes vital when ported to future technologies and test chips.

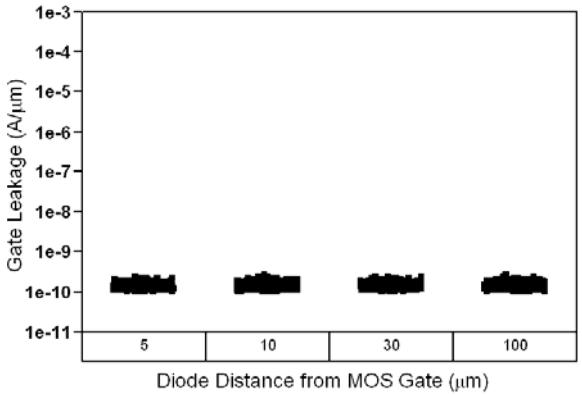



Fig. 8. Gate leakage distribution of varying diode distance from the MOS transistor. Each structure has a 5000:1 Metal1 to gate area antenna ratio.

## CONCLUSIONS

An experiment was developed to rigorously characterize the effects of plasma induced antenna gate leakage on the MOS transistor. The test structures were generated using a highly parameterized, hierarchical antenna template developed within an integrated system that automated test structure layout, placement, routing, and test plan development. The ability to efficiently design a more complete experiment makes it possible to identify issues earlier in the design cycle and optimize the final process monitor. The test structure generation procedure used to characterize plasma induced damage can be extended to other effects (e.g. electro-migration, matching) that impact early process development.

## REFERENCES

[1] V. Prabhakar, T. Brozek, X. Li, Y.D. Chan, C.R. Viswanathan, “Physical and Failure Analysis of Integrated Circuits”, 1995, 5th International Symposium, pp. 140-144.

[2] T. Hook, D. Harmon, C. Lin “Plasma process-induced damage on thick (6.8 nm) and thin (3.5 nm) gate oxide: parametric shifts, hot-carrier response, and dielectric integrity degradation” Microelectronics Reliability, Volume 41, Number 5, May 2001 , pp. 751-765(15).

[3] F. Shone, K. Wu, J. Shaw, E. Hokelek, S. Mittal, “Gate Oxide Charging and its Elimination for Metal Antenna Capacitor and Transistor in VLSI CMOS Double Layer Metal Technology,” Symposium VLSI Technology Digest, pp. 73-74, 1989.

[4] Z. Wang, A. Scarpa, S. Smits, C. Salm, F. Kuper, “Temperature Effect on Antenna Protection Strategy for Plasma-Process Induced Charging Damage,” International Symposium on Plasma and Process-Induced Damage, pp. 134-137, 2002.

[5] A. Gabrys, W. Greig, A. West, P. Lindorfer, W. French, S Mondal, et al “Integrated procedure automating test chip layout, place and route, and test plan development for efficient parametric device and process design,” in IEEE Transactions on Semiconductor Manufacturing, Vol. 22, No. 1, Feb. 2009.

[6] K. Cheung, P. Mason, J.T. Clemens, “Measuring plasma charging damage in ultra-thin gate-oxide," 5th International Symposium on Plasma Process-Induced Damage, pp. 10-13, 2000.

# Advanced Method for Measuring Ultra-Low Contact Resistivity Between Silicide and Silicon Based on Cross Bridge Kelvin Resistor

T. Isogai[1], H. Tanaka[2], A. Teramoto[2], T. Goto[2], S. Sugawa[1], and T. Ohmi[2, 3]

[1] Graduate School of Engineering, Tohoku University
[2] New Industry Creation Hatchery Center, Tohoku University
[3] WPI Research Center, Tohoku University

## ABSTRACT

In order to evaluate low contact resistivity precisely, we have developed a new test structure based on cross bridge Kelvin resistor. In this structure, the misalignment margin can be as small as possible. Furthermore, we had successively derived the theoretical expressions to ensure the validity of the newly developed method. This method will enable us to evaluate the silicide to silicon contact resistivity in the sub-$10^{-8}$ $\Omega cm^2$ region.

## INTRODUCTION

As the current drivability of MISFETs increases, the parasitic resistance in the source and drain region ($R_{SD}$) comes to limit their performance [1]. In order to reduce $R_{SD}$, it is important to reduce contact resistance. Because the contact area will inevitably be smaller, the only way to decrease contact resistance is to lower the contact resistivity. Although there have been reports that successfully obtained even in the $10^{-10}$ $\Omega cm^2$ range of contact resistivity of metal to metal or metal to silicide contact using cross bridge Kelvin resistor (CBKR) [2], [3], the measurement of ultra-low contact resistivity between silicide and silicon below $10^{-8}$ $\Omega cm^2$ region is still considered to be difficult, mainly due to the sheet resistance of heavily doped silicon, which is usually 1-2 orders of magnitude higher than that of metal or silicide. In this paper, a new test structure to detect ultra-low contact resistivity between silicide and silicon based on CBKR is proposed, and is applied to erbium silicide/$n^+$-Si contact.

## TEST STRUCTURE

Fig. 1 shows the simulation result of a typical relationship between extracted contact resistivity ($\rho_{Ce}$) and true contact resistivity ($\rho_C$) measured by CBKR with finite misalignment margin ($\delta$). These curves were calculated by using circuit simulator HSPICE, based on two-dimensional resistance network model [4]. As already reported by many authors, the extracted contact resistivity has lower limit mainly

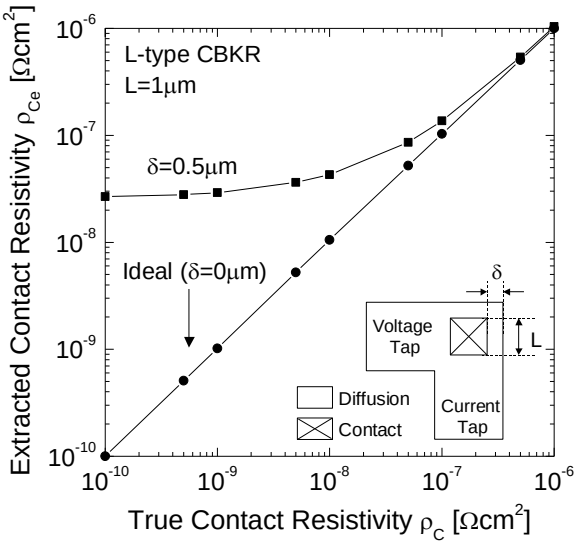



Fig. 1. A typical simulation result of the relationship between extracted and true specific contact resistivity with finite misalignment margin.

because of the lateral current crowding effect [5], [6]. One solution is to fabricate a very small contact, but it is difficult to fabricate sub-100-nm contact as it is originally designed. Another solution is to fabricate the CBKR with no $\delta$, also shown in Fig. 1 as the caption "ideal". Theoretically, it is possible to detect even $10^{-10}$ $\Omega cm^2$ by using 1-μm length contact when $\delta$ is zero. Therefore, one reasonable approach to measure low contact resistivity is to fabricate the CBKR with as small $\delta$ as possible.

Fig. 2 shows the concept of the advanced CBKR structure developed in this work. In order to fabricate the CBKR with very small $\delta$, we prepared 169 CBKRs with different alignment between diffusion layer (ACT) and contact layer (CNT). The misalignment margin is designed to be zero as shown in Fig. 2 (a). Because the misalignment can occur toward any directions, the position of CNT with reference to ACT is gradually shifted both x and y directions by $\pm n\Delta$ (n=1, 2, …6). Therefore even when misalignment occurs, certain sample out of 169 CBKRs should have least misalignment. In this work, the shift amount ($\Delta$) is designed to be 20 nm, so one sample out of 169

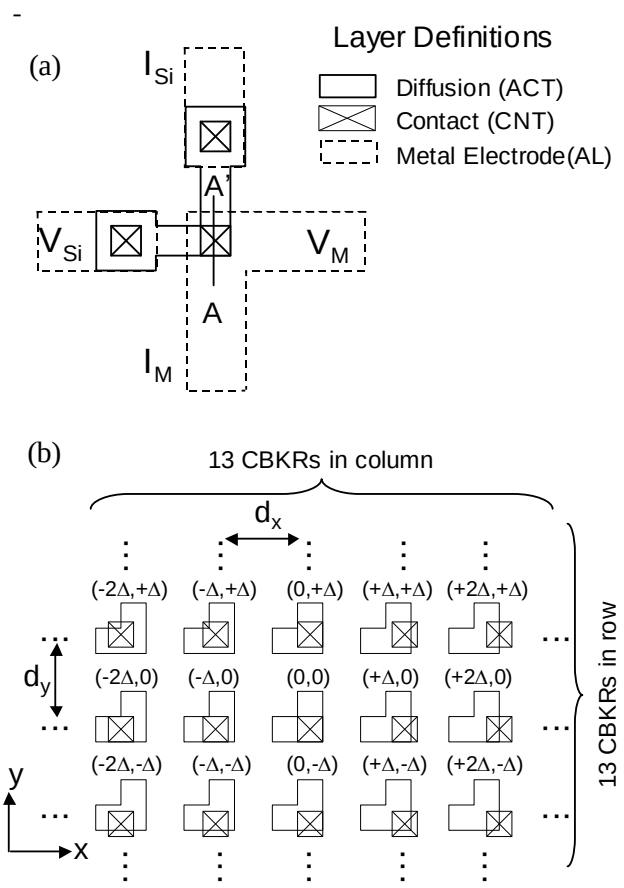



Fig. 2. (a) The structure of CBKR with no misalignment margin and its layer definitions. (b) The concept of CBKR array with gradually shifted contact position.

CBKRs should have misalignment of at most 10 nm to both x and y directions, provided the misalignment is smaller than ±6Δ (120nm). The shift amount can be smaller as long as the photomask is available. Because the CBKR has four 80 μm × 80 μm pads, the actual distance between each CBKR ($d_x$, $d_y$ in Fig. 2 (b)) is 320 μm.

## DEVICE FABRICATION

A 3-layer lithography process, using KrF (for ACT and CNT) and EB (for metal electrode: AL)was used for the device fabrication. A 100-nm-thick field oxide was formed on a p-type Si(100) wafer and the diffusion layer pattern was opened by buffer hydrofluoric (HF) acid. In order to investigate the effect of lateral dopant diffusion, two types of heavily doped region (A) shallower diffusion and (B) deeper diffusion were fabricated, as shown in Table 1. After a 100-nm-thick interlayer dielectric deposition, contact hole was opened. Then the activation anneal was performed, followed by silicidation process. Er (10 nm) with W (200 nm) capping layer was deposited, followed by 600 °C, 2 min annealing in Ar ambient to form erbium silicide. The carrier concentration at the silicide/silicon interface is ~$2\times10^{20}$ $cm^{-3}$ in both type

TABLE I
DIFFUSION LAYER FORMATION CONDITION

| | (A) | (B) |
|---|---|---|
| Ion Implantation | $As^+$, 25 keV, $2\times10^{15}$ $cm^{-2}$ | $As^+$, 50 keV, $6\times10^{15}$ $cm^{-2}$ |
| Activation Anneal ( in $N_2$) | 550°C, 60 min | 1,000°C, 10 min |
| Measured Sheet Resistance [Ω/□] | 141 | 33 |

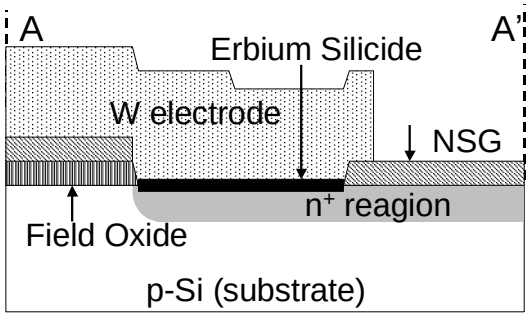



Fig. 3. A cross-sectional view of fabricated CBKR along the cutline A-A' in the Fig. 2 (a)

of devices. W capping layer is also used as a top electrode. Fig. 3 shows the cross-sectional structure of the fabricated CBKR. In order to avoid any plasma damages during $SiO_2$ etching, wet etching technique was employed. Therefore, relatively large contact (nominally 1 μm × 1 μm square) is mainly examined. The contact dimension was determined by measuring actual device by scanning electron spectroscope (SEM).

## RESULTS AND DISCUSSIONS

Fig. 4 (a) shows a schematic image of 13×13 array of CBKRs. In this figure, each square represents single CBKR device, shown in Fig. 4 (b). Fig. 4 (c) shows the SEM image around the contact hole of the fabricated device that has misalignments toward several directions before metalization process. It is found that CNT is exactly aligned to underlying ACT layer in the left image. The samples that have largest misalignment toward various directions (as shown arrows in images) are also shown in Fig. 4 (c). Although the best alignment position was (0, -2), marked as "A" in Fig.

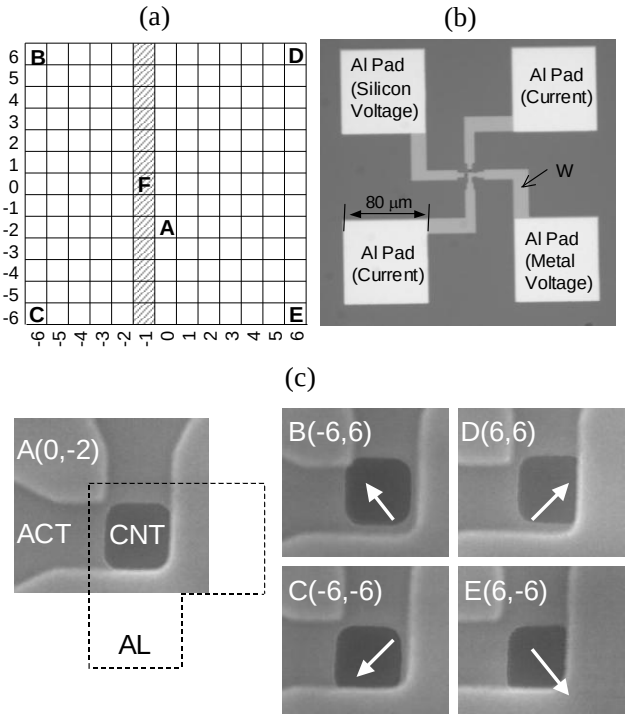



Fig. 4. (a) A schematic image of the CBKR array. (b)An optical Image of the fabricated device. (c) SEM images of fabricated devices around 1-μm contact hole for various misalignment directions.

4 (a) in this sample for SEM observation, the best position can vary for each sample because of the alignment variation of KrF stepper. For example, in the actually measured samples mentioned below, the best alignment position was (-1, 0), as marked "F" in Fig. 4 (a).

The contact resistivity of erbium silicide/$n^+$-silicon contact was measured by using this structure. The extracted value of contact resistivity (measured resistance × contact area) of the device having least misalignment are $6.3\times10^{-9}$ $\Omega cm^2$ for sample (A) and $2.7\times10^{-8}$ $\Omega cm^2$ for sample (B), respectively. The difference of the result can be explained by lateral dopant diffusion, which is discussed later. In principal, the CBKR structure developed in this work can be applied to measure much lower contact resistivity. However, the vertical voltage drop in the heavily doped Si region may limit the lower detectable $\rho_{Ce}$ values, as reported in the literature [6].

However, we must point out that the developed structure has an intrinsic problem. Because all samples have different alignment between ACT and CNT as described above, only 1 device out of 169 devices is available for the evaluation. This can be fatal because it is difficult to guarantee the accuracy and reproducibility of the extracted $\rho_{Ce}$. In order to solve this problem, we have derived the simple theoretical

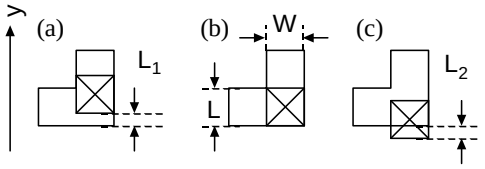



Fig. 5. The coordinate used for the theoretical calculation. Only the part of CBKR around the contact hole is shown.

expressions of $\rho_{Ce}$ when the contact position shifts to ± y direction (parallel to the current flow), and compared them to measured data. In this calculation, we neglect two-dimensional current flow. The coordinate used for the derivation is shown in Fig. 5. The calculation is based on 1-dimensional transmission line model [7] and is similar to those of reported in the literature [8].

(a) in the case CNT shifts to +y direction

$$\rho_{Ce} = \rho_C \frac{\gamma L_1 + \sinh\gamma(L - L_1)}{\sinh(\gamma L)} \quad (1)$$

(b) in the case CNT is exactly aligned to ACT

$$\rho_{Ce} = \rho_C \quad (2)$$

(c) in the case CNT shifts to –y direction

$$\rho_{Ce} = \rho_C \left\{1 + \frac{\gamma L_2}{\tanh\gamma(L - L_2)} + \frac{(\gamma L_2)^2}{2}\right\} \quad (3)$$

From (1), $\rho_{Ce}$ becomes smaller than $\rho_C$, because although contact area is unchanged in this case, voltage tap does not detect whole voltage drop in the contact. On the contrary, from (3), $\rho_{Ce}$ becomes larger than $\rho_C$, because in this case contact area becomes smaller and less current can flow. Note also that both (1) and (3) converge to $\rho_C$ when $L_1$ and $L_2$ approach to zero, which ensures the validity of these expressions. Fig. 6 shows the comparison between measured data (from 13 devices centered by "F" as marked in Fig. 4 (a)) and theoretical curves using expressions (1) and (3). Although the measured data show some fluctuation in sample (A), they can be well fitted by there expressions. The true contact resistivity in this case should be in the range of $5.0\text{-}8.3\times10^{-9}$ $\Omega cm^2$. In sample (B), because a large (over 0.1μm) amount of lateral dopant diffusion occurs, the CBKR structure effectively becomes L-type like. Therefore applying above expressions may not be appropriate. However, also in this case, the experimental data are well fitted by these expressions, and the extracted value is 2.6-

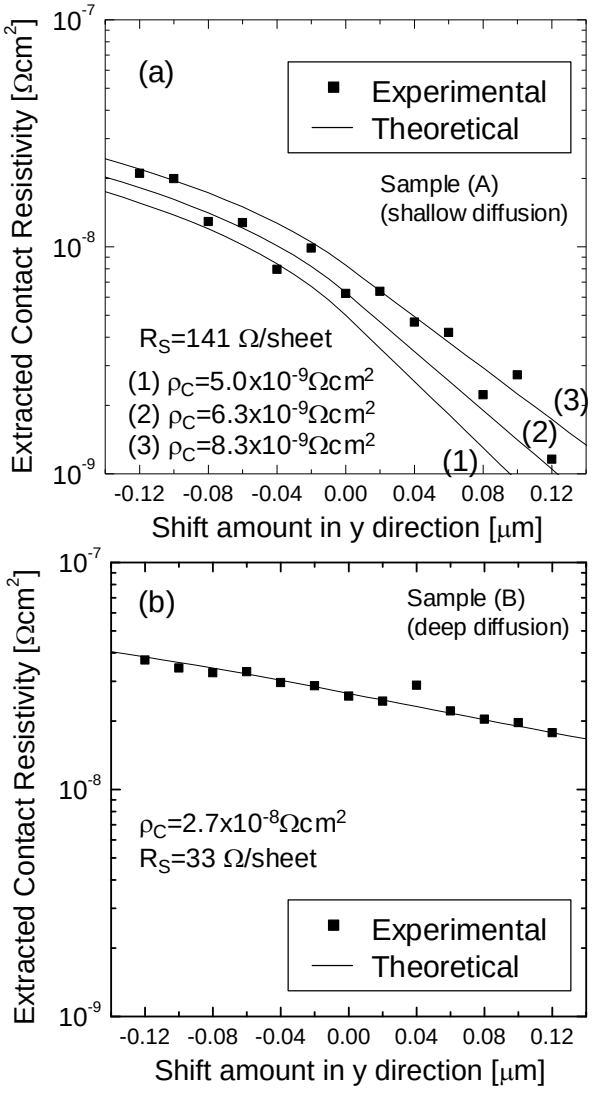



Fig. 6. The measured data and theoretical fitting results for both samples.

$2.8\times10^{-8}$ Ωcm$^2$ in sample (B). Therefore, the $\rho_{Ce}$ value obtained in this work is considered to be reasonable from at least 13 measured points, not from just 1 point. By using this method, we could directly extract ultra-low contact resistivity. Fig. 7 shows the misalignment margin dependence of the extracted contact resistivity, using both the new structure and a conventional L-type CBKR. When the misalignment margin is relatively large, the CBKR having deeper diffusion layer gives smaller extracted value because of its low sheet resistance of Si layer. However, when the misalignment margin becomes smaller enough, which is realized by the structure developed in this work, the CBKR with deeper diffusion cannot detect the precise measurement value because of the lateral dopant diffusion. On the contrary, the CBKR having shallower diffusion layer gives much lower value of $\rho_{Ce}$ due to little lateral diffusion, while such device gives much larger values with the increased misalignment margin because of its high sheet resistance. Therefore, the structure developed in this

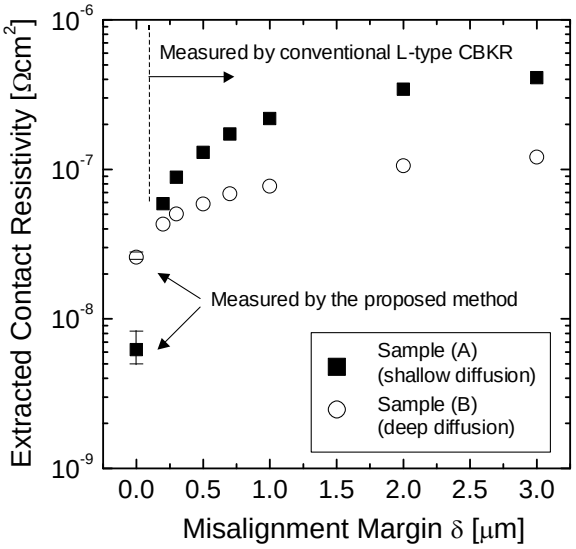



Fig. 7. Misalignment margin dependence of extracted contact resistivity for both samples. Measured results by proposed method are also shown.

work that has least misalignment margin, as well as shallower diffusion layer is a useful way to detect ultra-low contact resistivity between silicide and silicon below $10^{-8}$ Ωcm$^2$ range. Although low-temperature annealing method [9] was employed to form shallow diffusion layer in this work, an optimized rapid thermal annealing (RTA) can be also applied.

## CONCLUSION

We have developed the advanced method to evaluate ultra-low contact resistivity between silicide and heavily doped silicon and applied it to erbium silicide/n$^+$-silicon contact. The proposed method enables us to fabricate the CBKR with very small misalignment and therefore precisely detect the low contact resistivity below $10^{-8}$ Ωcm$^2$ region. We also derived simple theoretical expression to ensure its validity. The technology given in this paper is useful to develop low-resistance silicide/silicon contact in the source/drain region and will help us to realize high performance MOS devices.

## REFERENCES

[1] T. Ohmi, A. Teramoto, R. Kuroda, N. Miyamoto, IEEE Transactions on Electron Devices Vol. 54, No. 6, June, 2007, pp. 1471-1477.

[2] M. Bhaskaran, S. Sriram, A. S. Holland, IEEE Electron Device Letters, Vol. 29, No. 3, pp. 259-261, March, 2008.

[3] N. Stavitski J. H. Klootwijk, H. W. van Xeijl, A. Y. Kovalgin, R. A. M. Wolters, in Proc. IEEE ICMTS, Edinburgh, UK, pp. 199-204, 2008.

[4] M. Finetti, A. Scorzoni, G. Soncini, IEEE Electron Device Letters, Vol. EDL-5, No. 12, pp. 524-526, December, 1984.

[5] J. Santander, M. Lozano, C. Cane, IEEE Transactions on Electron Devices, Vol. 40, No. 5, pp. 944-950, May, 1993.
[6] S. Holland, G. K. Reeves, P. W. Leech, IEEE Transactions on Electron Devices Vol. 51, No. 6, pp. 914-919, June 2004.
[7] H. H. Berger, Solid-State Electronics, Vol. 15, p. 145-158, 1972.
[8] M. Ono, A. Nishiyama, A. Toriumi, Solid-State Electronics, Vol. 46, pp. 1325-1331, 2002.
[9] T. Nitta, T. Ohmi, Y. Ishihara, A. Okita, T. Shibata, J. Sugiura, N. Ohkawa, Journal of Applied Phsics, Vol. 67, No. 12, pp. 7404-7412, June, 1990.

# A Test Structure for Statistical Evaluation of Characteristics Variability in a Very Large Number of MOSFETs

S. Watabe[1], S. Sugawa[1], K. Abe[1], T. Fujisawa[1], N. Miyamoto[2], A. Teramoto[2], and T. Ohmi[3]

[1] Graduate School of Engineering, Tohoku University
[2]New Industry Creation Hatchery Center, Tohoku University
[3]WPI Research center, Tohoku University
Aza-Aoba 6-6-10, Aramaki, Aoba-ku, Sendai 980-8579, Japan

## Abstract

We have proposed and developed a test structure for evaluating electrical characteristics variability of a large number of MOSFETs in very short time using very simple circuit structure. The electrical characteristics such as threshold voltage, subthreshold swings (S-factors, random telegraph signal noise, and so on, can be measured in over one million MOSFETs. This new test structure circuit and results measured by this circuit are very efficient in developing processes, process equipment and device structure which suppress variability.

## Introduction

The increase of electrical characteristic variability in MOSFETs caused by miniaturization of MOSFETs is one of critical issues for realizing the low power consumption for LSIs and high accuracy for analog devices [1-3]. The development and improvement of fabrication processes, process equipments, and device structures for reducing characteristics variability are very important. It is necessary to evaluate easily the variability of very large number of MOSFETs in very short time. In this paper, we propose a new test structure for statistical evaluation. The electrical properties such as $I_{ds}$-$V_{gs}$ characteristics, and threshold voltages, subthreshold swings, random telegraph signal noise, and so on can be measured by this structure for one million MOSFETs in very short time. Actually about one million MOSFETs can be measured in 0.7 sec for one bias condition. This test structure is efficient for development and improvement of processes, process equipments, and device structures.

## Test Structure

Fig. 1 shows the circuit schematic view of newly developed test structure for measuring electrical characteristics in a very large number of MOSFETs in short time. The test structure is simply composed of the arrayed MOSFETs included the measured and select switch transistors (A) and simple peripheral circuits, which are the vertical and horizontal shift registers (B, C) to address measured transistors, current source transistors located on every column for current control of the measured transistors (D), analog memories located on every column to store source voltages of measured MOSFETs (E), and source follower circuit to amplifies output voltages (F). It is easy to integrate many MOSFETs in a very small area because the unit cell is constructed with only two select and measured MOSFETs. Because this test structure employs very simple circuits, it can operate in a wide range, as a result, it can evaluate various MOSFETs which have different gate lengths, gate widths, gate insulators, its thicknesses, and so on. Fig. 2 shows the circuit schematic view of a unit cell, constructed with measured MOSFET and select transistor, and current source transistor in Fig. 1, which indicates the principle of electrical characteristics measurement. Measured transistor and current source transistor construct the source follower circuit through the select switch transistor. When the gate voltage ($\Phi_x$) of the select transistor is biased to high level signal according to the vertical shift register circuit, the constant current ($I_{REF}$) controlled by applying voltage ($V_{REF}$) flows to the measured MOSFET. $I_{REF}$ is the same as $I_{ds}$ in measured

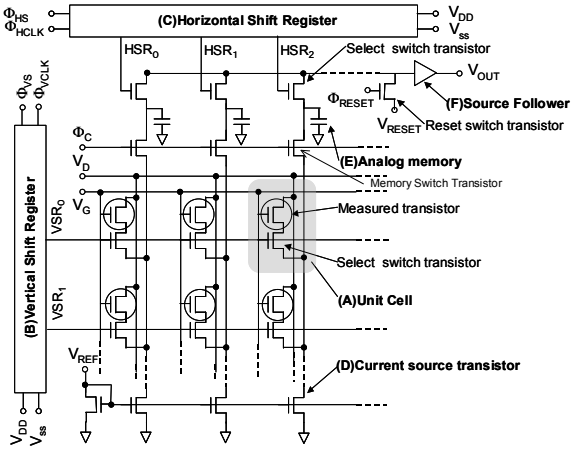



Fig. 1 Circuit schematic view of the newly developed test structure, which is simply composed of the MOSFETs array included the measured and select switch MOSFET (A), vertical and horizontal shift register to address measured MOSFETs (B, C), current source transistors to determine the drain current of the measured MOSFET (D), analog memory to store output voltages (E), source follower circuit to amplify output voltages (F). Drain voltage $V_D$, gate voltage $V_G$ in measured MOSFETs are applied from external voltage sources simultaneously.

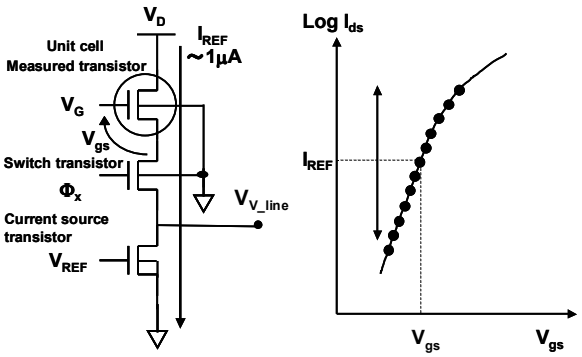



Fig. 2 Schematic view of unit cell and measurement method of $I_{ds}$-$V_{gs}$ characteristics. It shows the principle of measurement of electrical characteristics. Measured transistor and current source transistor construct the source follower circuit through the select switch transistor. $I_{ds}$-$V_{gs}$ characteristics are measured by changing $I_{REF}$, which is controlled by $V_{REF}$. In this experiment, the dimensions of switch transistor and current source transistor are L = 0.34 μm, W = 1.26 μm, and L = 10 μm, W = 15 μm, respectively.

transistor. Because $I_{REF}$ is independent of the voltage between drain and source, which is the same as the output voltage ($V_{v\text{-}line}$), in current source transistor, because the current source transistors operate at saturation region. Then $V_{v\text{-}line}$ is shown as following equation:

$$V_{gs} = V_G - V_{v-line} - I_{ds} \cdot R_{select}, \qquad (1)$$

where $V_G$ is the applied gate voltage, $V_{gs}$ is the voltage between the gate and source of the measured MOSFET and $R_{select}$ is the channel resistance of the select switch transistor. When the select switch transistor with W/L of 1.26 μm / 0.34 μm turns on, $R_{select}$ is approximately 400 ohm. When $I_{REF}$ is 1 μA, I*R voltage drop in select switch transistor is approximately 0.4 mV, and it's enough small. The threshold voltage and size variation of the select transistor contributes to only the variation of I*R voltage drop, and does not influence the measurement error. Equation (1) is shown as (2) approximately:

$$V_{gs} \approx V_G - V_{v-line}. \qquad (2)$$

We can measure $V_{v\text{-}line}$-$V_{out}$ characteristic by changing $V_D$ when the measured tranistors turn on. Fig. 3 shows $V_{v\text{-}line}$-$V_{out}$ characteristic in this experiment. The test circuit has good linearity in the range of $V_{v\text{-}line}$ from 0.8 V to 2.3 V. Therefore, we can obtain $V_{gs}$ of measured transistor when $I_{REF}$ flow to it by the measuring of $V_{out}$ according to (2) and Fig. 3. Furthermore, we can measure $I_{ds}$-$V_{gs}$ characteristics of measured transistors by changing $I_{REF}$, which is controlled by $V_{REF}$.

The variability can be suppressed by enlarging device size. Then, the gate size of current source transistors should be much larger than that of measured transistors to prevent the variability of $I_{REF}$ in each vertical line [3-5]. In this experiment, the gate size of

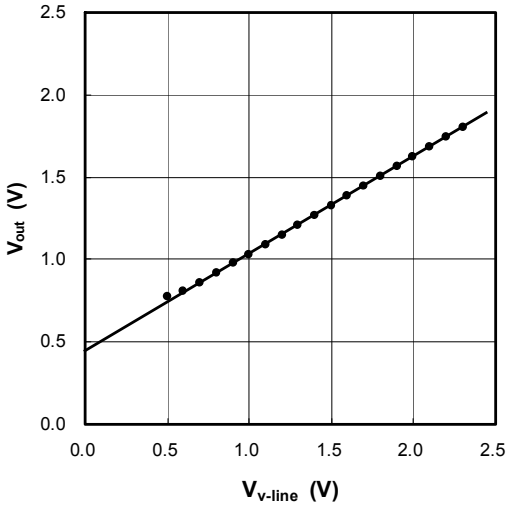



Fig. 3 $V_{v\text{-}line}$-$V_{out}$ characteristic of the test structure at $V_{DD}$ = 3.3 V, $V_{reset}$ = 1.2 V in Fig. 1 in this experiment. This test circuit has good linearity in the range of Vv-line from 0.8 V to 2.3 V.

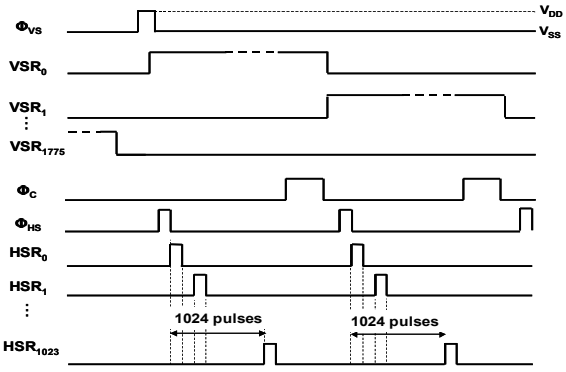



Fig. 4 The timing diagram of operation for test structure. When a start pulse of the vertical shift resistor ($\Phi_{VS}$) is applied externally, each select transistor turns on one by one. When the memory switch transistors turn on by applying ($\Phi_C$), the output signals ($V'_{out}$) in one line store in the analog memory (E). Next, a memory switch transistor turns off, then, start pulse ($\Phi_{HS}$) of the horizontal shift resistor is applied, the signals in one line are sent by amplification at the source follower circuit. During this read-out period, the next line is selected by the vertical shift resistor and then the output signal is charging the parasitic capacitance of the wiring.

current source is MOSFETs with W/L of 15 μm / 10 μm.

The routine to measure is described as follows. Fig. 4 shows the timing diagram of operation for this test structure. When a start pulse of the vertical shift resistor ($\Phi_{VS}$ in Fig. 1) is applied externally, each select transistor turns on one by one. When the memory switch transistor turns on by applying $\Phi_C$, the output signals ($V_{v\text{-}line}$) in one line store in the analog memory (E). Next, a memory switch transistor turns off, then, start pulse ($\Phi_{HS}$) of the horizontal shift resistor is applied, the signals in one line are sent for amplification at the source follower circuit. During this read-out period, the next line is selected by the vertical shift resistor and then the output signal is

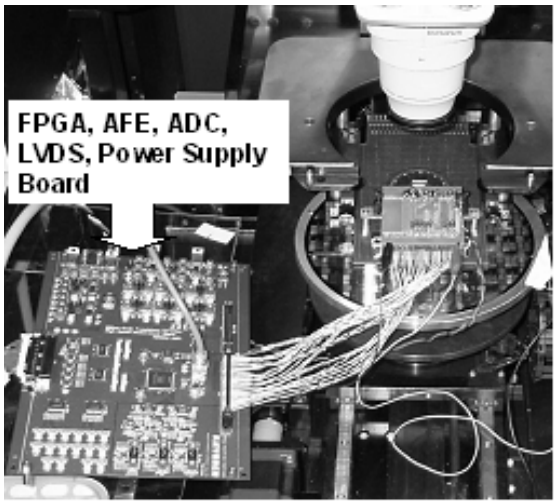



Fig. 5 The photograph of measurement system.

charging the parasitic capacitance of the wiring. The photograph of measurement system is shown in Fig. 5. The output signal is amplified as voltage signal in the chip and the analog output signals are transformed to a digital signal by the A/D converter set near the chip. In this experiment, about one million MOSFETs can be measured in 0.7 sec for one bias condition. In this test structure the signals are treated as current signals only in the vertical wiring and the analog memory, those capacitances have about 10 pF to reduce the thermal noise to less than 0.06 mV. It is possible to measure in short time because the charging time of the vertical wiring and the analog memory by the small current $I_{ref}$ can be secured enough the vertical shift register pulse is turn on, and because the read-out of output signals and the charging of parasitic capacitance by the next output signal can be operated in parallel by separating between cell region and read-out circuit.

In this test structure, we also identify easily some MOSFETs having random telegraph signal noise by time scale measurement [6].

### Measurements, Results and Discussion

The test structure is manufactured by 0.22 μm, 1-poly 2-metal standard CMOS technology and includes MOSFETs of several gate sizes as shown in table 1. The gate insulator is pyrogenic oxidation and its thickness is 5.8 nm.

Fig. 6 shows the layout pattern of a unit cell with and without antenna for accelerating the plasma damage. The antenna is formed as finger structure with line and space of 0.29 μm and 0.29 μm in first metal layer, and is connected to the gate electrode of measured MOSFET with W/L of 0.24 μm/ 0.30 μm. First metal thickness is 600 nm. The antenna ratios of upper area are 23, 100, 1000, and 10000, in including side area are 40, 420, 4100 and 41000.

The output voltages $V_{out}$ for one transistor in 300 times measurement are shown in Fig. 7. The variation in

Table 1 Measured transistor dimensions and numbers in test structure.

| Gate length L (μm) | Gate width W(μm) | Number of transistors |
|---|---|---|
| 0.22 | 0.28 | 65,536 (64x1024) |
| 0.22 | 0.30 | 65,536 (64x1024) |
| 0.24 | 0.30 | 65,536 (64x1024) |
| 0.24 | 1.5 | 65,536 (64x1024) |
| 0.24 | 15 | 16,384 (64x256) |
| 0.40 | 15 | 16,384 (64x256) |
| 0.40 | 1.5 | 65,536 (64x1024) |
| 1.2 | 0.30 | 32,768 (32x1024) |
| 1.2 | 1.5 | 32,768 (32x1024) |
| 4.0 | 0.30 | 32,768 (32x1024) |
| 4.0 | 1.5 | 32,768 (32x1024) |
| 0.24 | 0.30 | 32,768 (32x1024) Antenna ratio 100 |
| 1.2 | 15 | 8,192 (32x256) |
| 4.0 | 15 | 8,192 (32x256) |
| 10 | 15 | 4,096 (16x256) |
| 0.24 | 0.30 | 4,096 (16x256) Antenna ratio 1,000 |
| 0.24 | 0.30 | 1,344 (32x42) Antenna ration 10,000 |

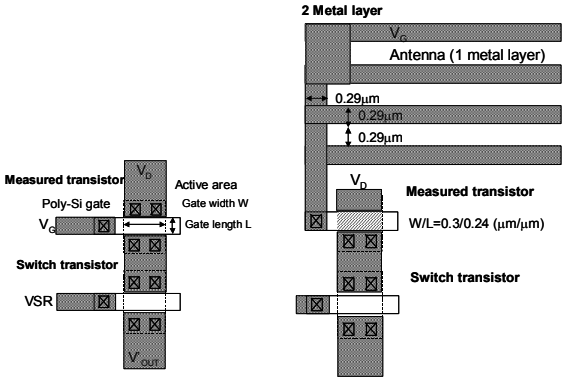



Fig. 6 Layout pattern of measured transistors with and without antenna. The antenna is put in the transistors with gate length of 0.24 μm, gate width of 0.30 μm and finger type structure with line and space of 0.29 μm and 0.29 μm in first Metal layer. Antenna ratio of transistors without antenna is 23.

sampling number that indicates the measurement accuracy is 1.9mV as 3σ.

$I_{ds}$-$V_{gs}$ characteristics of nMOSFET and pMOSFET with W/L of 0.30 μm / 0.24 μm are shown in Figs. 8. The range of $I_{ds}$ is from 0.5 nA to 5 μA. In this case, the average substrate bias ($V_{bs}$) and the voltage between source and drain ($V_{ds}$) are -1 V and 1.5 V for nMOSFETs, and, 1.3 V and -1.2 V for pMOSFETs, respectively. The $V_{gs}$ variation (Max.–Min.) of 65,536 nMOSFETs for each bias point reaches about 200 mV.

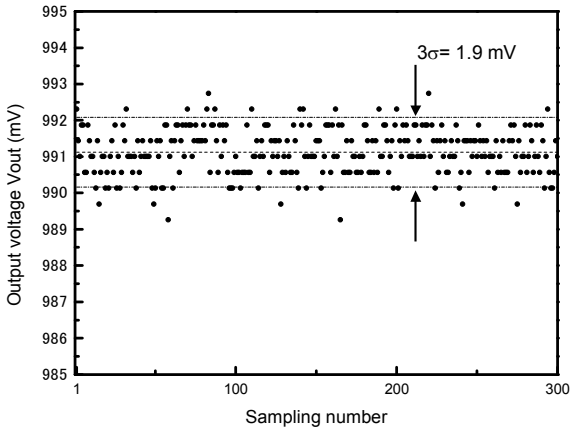
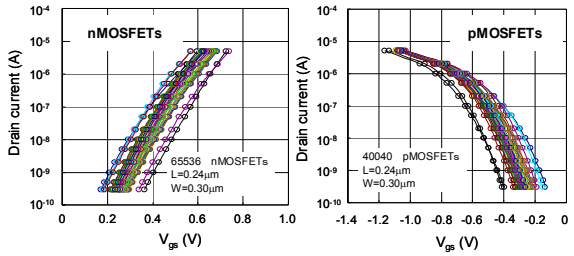



Fig. 7 The measurement accuracy for the output voltage $V_{out}$ in one cell.
The variation in sampling number is 1.9 mV as 3σ.

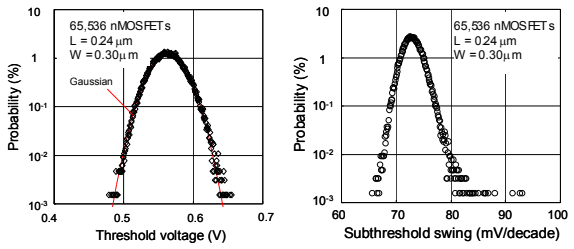
Figs. 9 show threshold voltage ($V_{th}$) and subthreshold swing (S-factor) distribution. The $V_{th}$ is defined as $V_{gs}$ at $I_{ds}$ of 1 μA. S-factor is defined as follows:

$$S = \frac{d(\log I_{ds})}{dV_{gs}} (@ I_{ds} = 10^{-9} A) \quad . \qquad (3)$$

In Figs. 9 the solid line shows the Gaussian distribution fitted to the measurement data. The distribution fits the Gaussian distribution well, and its standard deviation $\sigma(V_{th})$ is 20.9 mV. S-factors of nMOSFETs distribute from 65 to 93 mV/decade. We can evaluate the variation of $V_{th}$ and S-factor in short time by using the developed test pattern.
Fig. 10 shows $\sigma(V_{th})$ as a function of $(LW)^{-1/2}$. $\sigma(V_{th})$ is almost proportional to $(LW)^{-1/2}$ as reported in [3-5]. However some $\sigma(V_{th})$ of MOSFETs with gate sizes are not on line in Fig. 10.
Fig. 11 shows the threshold voltage map in 200 mm wafer. $V_{th}$ variability in a wafer is visible by being displayed as a color scale. The blue and red show high and low $V_{th}$, respectively. In a wafer, we can observe that $V_{th}$ in center region is higher than that in the peripheral region of the wafer. This map is very useful for development and improvement of processes and process equipments.
Figs. 12 show $I_{ds}$-$V_{gs}$ characteristics for nMOSFETs with W/L of 0.30 μm / 0.24 μm having antenna ratios of 23, 1,000 and 10,000 by measured developed test structure and with the same gate area measured by conventional test structure with four terminal pads per one MOSFET at five points in a wafer, shown in solid line. Gate terminal pad becomes antenna, which is square type, and does not have the protection diode.




Figs. 8 Ids-Vgs characteristics of nMOSFETs and pMOSFETs.




Figs. 9 Threshold voltages and subthreshold slopes (S-Factor) distribution of 65,536 MOSFETs.

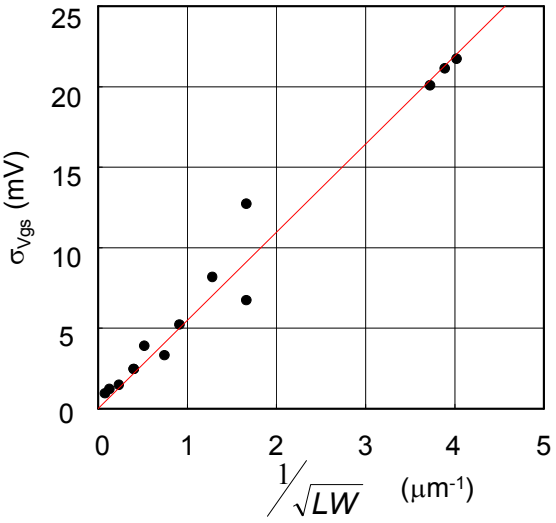



Fig. 10 Standard deviations of $V_{th}$ as a function of $(LW)^{-1/2}$.
$\sigma(V_{th})$ is almost proportional to $(LW)^{-1/2}$. Some $\sigma(V_{th})$ of MOSFETs with gate sizes is not on line. The proportionality factor is 5.8 mV·μm.

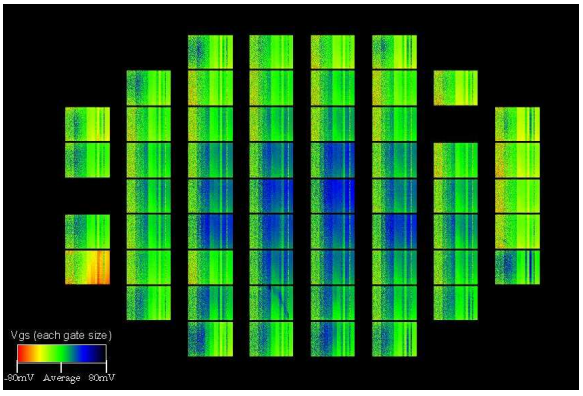



Fig. 11 The threshold voltage map in 200 mm wafer. $V_{th}$ variability in a wafer is visible by being displayed as a color scale. The blue and red show high and low $V_{th}$, respectively.
In a wafer, we can observe that $V_{th}$ in center region is higher than that in the peripheral region of the wafer. Some blank chips in peripheral region show that we cannot measure because the chips were broken. This map is very useful for development and improvement of processes and process equipment.

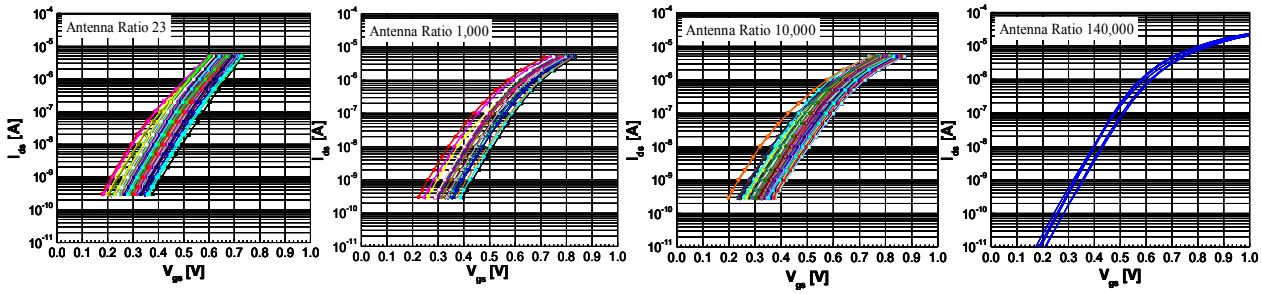



Fig. 12 $I_{ds}$-$V_{gs}$ characteristics for nMOSFETs with W/L of 0.30 μm / 0.24 μm having antenna ratios of 23, 1,000, and 10,000 by measured developed test structure and with same gate area measured by conventional test structure with four terminal pads per one MOSFET at five points in a wafer. Gate terminal pad becomes antenna, which is square type. The antenna ratio of MOSFETs in conventional test structure is about 140,000. The slopes of $I_{ds}$-$V_{gs}$ characteristics decrease with an increase in antenna ratio.

The antenna ratio of nMOSFETs in conventional test structure is about 140,000. The slopes of $I_{ds}$-$V_{gs}$ characteristics decrease with an increase in antenna ratio. It is considered that charge-up occurs in the plasma processes after antenna formation. It indicates that MOSFETs in LSI have the large variability of characteristics because of having various antenna ratios in LSI. Characteristics for MOSFETs of conventional test structure almost correspond to those for MOSFETs with antenna ratio of 10,000 in developed test structure. In conventional test structure, the density of measured MOSFETs is about one per 100-μm-square because of the area of terminal pads, and the measurement of the large number of MOSFETs takes long time. Then, it is difficult for the conventional test structure to evaluate the characteristics variability occurred in LSI statistically. The developed test structure is efficient for the development and improvement of process, process equipment, and device structure included the evaluation of characteristics variability because it enables the rapid measurement very easily.

## Conclusions

We have proposed and demonstrated a new test structure for statistical characterization of a very large number of MOSFETs in very short time, actually measuring one million MOSFETs takes 0.7 seconds for one bias condition. The test structure can rapidly measure $I_{ds}$-$V_{gs}$ characteristics and extract easily threshold voltages, subthreshold swings and so on. The rapid measurement by using this test structure enables to evaluate and analyze variability of electrical property in MOSFETs not only in chip also in wafer, lot and production line.

The developed test structure is efficient for out design, processes, process equipment and device development to suppress the variability of electrical property in MOSFETs for realizing low power consumption of ULSI and high accurate analog circuit.

## Acknowledgment

We would like to thank Dr. M. Toita, Mr. Y. Kamata and Mr. K. Shibusawa for useful discussions and manufacture of the test structure.

## References

[1] B. Hoeneisen and C. A. Mead, “Fundamental Limitations in Microelectronics-I. MOS Technology,” Solid State Electron., vol. 15, pp. 819-829, 1972.

[2] T. Ohmi, M. Hirayama, and A. Teramoto, “New era of silicon technologies due to radical reaction based semiconductor manufacturing,” J. Phys. D, Appl. Phys., vol. 39, pp. R1-R17, 2006.

[3] M. Pelgrom, A. Duinmaijer, and A. Welbers, “Matching properties of MOS transistor,” IEEE J. Solid-State Circuits, vol. 24, pp. 1433-1440, 1989.

[4] S. T. Ma, A. Keshavarzi, V. De, and J. R. Brews, “A Statistical Model for Extracting Geometric Sources of Transistor Performance Variation,” IEEE Trans. Electron Devices, vol. 51, pp. 36-41, 2004.

[5] A. Stolk, F. P. Widdershoven, and D. B. M. Klaassen, “Modeling Statistical Dopant Fluctuations in MOS Transistors,” IEEE Trans. Electron Devices, vol. 45, pp. 1960-1971, 1998.

[6] K. Abe, S. Sugawa, S. Watabe, N. Miyamoto, A. Teramoto, Y. Kamata, K. Shibusawa, M. Toita, and T. Ohmi, “Random Telegraph Signal Statistical Analysis using a Very Large-scale Array TEG with 1M MOSFETs,” in Symp. VLSI Tech. Dig., pp. 210-211, 2007.

# Estimating MOSFET Leakage from Low-cost, Low-resolution Fast Parametric Test



Sharad Saxena, Takumi Uezono, Rakesh Vallishayee[2], Ron Lindley[1], Andrew Swimmer[1], Steve Winters[1]

PDF Solutions Inc., Richardson, TX 75082, USA
Phone: +1-972-889-3085, FAX: +1-972-889-2486, Email: sharad.saxena@pdf.com
[1]PDF Solutions Inc., San Diego, CA 92128, USA
[2]PDF Solutions Inc., San Jose, CA 95110, USA

## ABSTRACT

A method of estimating the subthershold component of MOSFET off-state current ($I_{offs}$) using low-cost, low-resolution fast parallel parametric test is introduced. This method measures the subthreshold slope and uses it to estimate $I_{offs}$. Measurements of individual transistors show a very good agreement between measured $I_{offs}$ and $I_{offs}$ estimated using our approach. For a simple pad-efficient transistor array test-structure, where unselected devices can add additional noise to the subthreshold measurements, the sum of extracted $I_{offs}$ for all transistors in an array is strongly correlated to the measured array $I_{offs}$, even though it does not match the measured array $I_{offs}$. The strong correlation is used to derive calibration factors which are then used to estimate individual transistor $I_{offs}$ from array test structures. This allows statistical characterization of transistor leakage during volume production with minimal test time overhead. The applications of statistical off-state leakage characterization to diagnose IDDQ yield problems during production are also described.

## INTRODUCTION

Leakage and variation are two undesirable consequences of transistor scaling. IDDQ is a critical parameter for most ICs, and ensuring high IDDQ yield requires characterizing not just the average values of the various components of leakage, but also the distribution of the leakage components.

The different sources of variation: systematic layout driven, lot-to-lot, wafer-to-wafer, die-to-die, across-chip and local variation all impact transistor leakage. Characterization and modeling the impact of these different sources of variation on leakage requires a large sample of measurements to separate the contribution of each source of variation on the different components of leakage.

Leakage monitoring during volume production requires fast measurements to minimize the test-time overhead. However, fast measurement of leakage poses special problems. Firstly, leakage currents are small, which implies that high-resolution measurements units are required. The expense of these high-accuracy units makes the use of parallel measurements to reduce test time cost-prohibitive. In contrast, moderate and strong inversion device characteristics like drive current (Idrive), threshold voltage (Vth) and transconductance (gm) can be accurately measured with low-resolution units allowing fast parallel test [1][2]. Secondly, accurate measurement requires long integration times.

This paper describes a method for fast estimation of transistor subthreshold current in off-state ($I_{offs}$) using low resolution measurement units. Gate induced drain leakage (GIDL) and gate current are the other main components of leakage in nanometer scale technologies. The technique of this paper does not allow the estimation of GIDL and gate current. They are estimated using separate test structures [2].

The $I_{offs}$ estimation method makes use of $I_d$-$V_g$ curves measured in sub-threshold. The sub-threshold currents are relatively larger than $I_{offs}$, allowing the use of low-resolution measurement units. Low-resolution measurements units in turn enable cost-effective parallel testing. When combined with our method for fast measurement of other devices characteristics (Idrive, Vth, gm, etc) [1][2], the $I_{offs}$ estimation does not add any additional measurement overhead, it reuses the same I-V curves used to measure or extract the other parameters.

## LEAKAGE CURRENT ESTIMATION METHOD

Drain-source current $I_{ds}$ in subthreshold region is modeled by following expression assuming $V_{ds}$ is larger than the thermal voltage [3]

$$I_{ds} = I_0 10^{(Vgs-Vt)/SS},$$

where $I_0$ is drain-source current when $V_{gs} = V_t$, $V_{gs}$ is gate-source voltage, $V_t$ is threshold voltage, and SS is subthreshold slope. From this equation, the plot of $\log(I_{ds})$ vs. $V_{gs}$ is expected to be linear in sub-threshold region with a constant slope. By extrapolating $\log(I_{ds})$ vs. $V_{gs}$ plot from the point where SS calculated, $I_{offs}$ can be estimated as $\log(I_{ds})$-intercept. $I_{offs}$ estimation method for nMOS can be summarized as follows:

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

1. measure $I_{ds}$ vs. $V_{gs}$ curve for Vgs > 0;
2. calculate subthreshold slope (SS) by linear regression on consecutive sets of three $\log(I_{ds})$-$V_{gs}$ points;
3. extrapolate the linear regression to $V_{gs}$=0

Two modifications are made to the above procedure to account for the use of low-resolution measurements units and GIDL. In the first modification the SS is calculated from points where $I_{ds} > I_{res}$, where $I_{res}$ is the resolution of the equipment. This has the additional benefit that when this technique is used to estimate leakage of transistors in a device array, the impact of the leakage of the non-selected devices in device arrays is minimized [1][4][5][6]. The second modification is to check for quality of fit ($R^2$) of the linear regression. If the $R^2$ is less than a threshold, the Id-Vg curve is rejected and $I_{offs}$ is not extracted for such an abnormal device. An example of an abnormal device is a transistor that turns on before $I_{res}$ is reached. Devices are also rejected if the slope of the regression is negative. Negative slope indicates devices with very large GIDL currents. These modifications allow a robust estimation of $I_{offs}$ from low resolution measurement units.

## EXAMPLES

The accuracy of the proposed method was assessed using measurements obtained from transistors fabricated in a 45 nm bulk CMOS process. First, the extrapolation method was applied to individual devices whose pins are directly connected to pads and are not shared with other devices. Figure 1 shows the comparison of $I_{offs}$ measured directly on these structures and $I_{offs}$ extracted from SS for a 45nm CMOS technology. As can be seen, a very good match is obtained between directly the measured $I_{offs}$ and the $I_{offs}$ estimated from SS. The correlation between extrapolated off current and measured off current is greater than 0.99 and the error of almost all of devices is less than 10%.

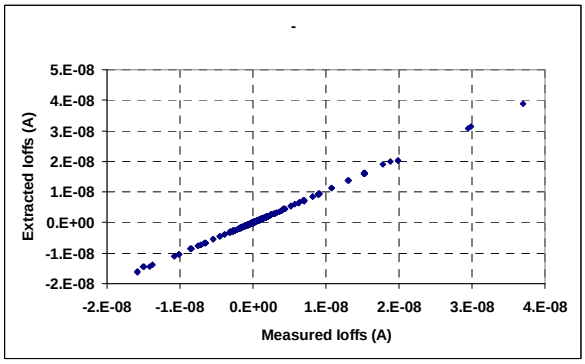



Figure 1: Comparison of $I_{offs}$ estimated from SS extrapolation with direct measurement. The plot includes both nMOS and pMOS.

### Transistor Arrays

Transistor arrays are compact and pad-efficient test structures for characterization of transistor variability [1][4][5][6]. This paper reports the results of applying $I_{offs}$ estimation to a simple device array, which has been placed on scribe line of product wafers [1][2]. Figure 2 is a schematic of this test structure. In this transistor array 32 transistors share the source and drain pins. The source and drain pins are directly connected to the pads. Several arrays are implemented in one pad group to be tested in parallel. Figure 3 shows the distribution of $I_{offs}$ extracted from parallel measurements made during volume production on device arrays placed on scribe line. The expected log-normal distribution of $I_{offs}$ is observed.

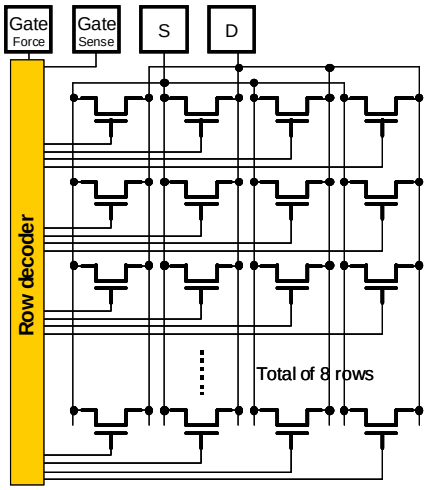



Figure 2: Schematic of transistor array test structure.

In addition to moderate and strong inversions characteristics of individual transistors in an array (eg Idrive, Vth, gm), this structure provides the measurement of $I_{off}$ of the complete array on the source side. Array $I_{offs}$ is the sum of the off-state current of all transistors in the array and is obtained by measuring the current on the source pad when no transistor is selected.

Figure 4 and Figure 5 shows the comparison of Array $I_{offs}$ to the sum of extracted $I_{offs}$ of the transistors in an array. Correlation is greater than 0.9 for pMOS and nMOS, wide and narrow transistors. However, the sum of extracted $I_{offs}$ does not match the array $I_{offs}$. The likely cause of this discrepancy is the contribution of the non-selected devices in the array to the measured $I_{ds}$. Alternate device array architectures that minimize the contribution of non-selected devices to the measured $I_{ds}$ could improve this match [4][5][6]. However, these techniques often require extra pads, which limit the number of arrays that can be put in a pad limited applications like scribe lines.

The high value of the correlation coefficient allows the extracted $I_{offs}$ to be a good indicator of $I_{offs}$, for example to compare relative values across wafers and

across different splits. Moreover, the strong correlation makes it possible to estimate calibration factors which are used to calibrate the extracted $I_{offs}$. The next section describes the calibration procedure.

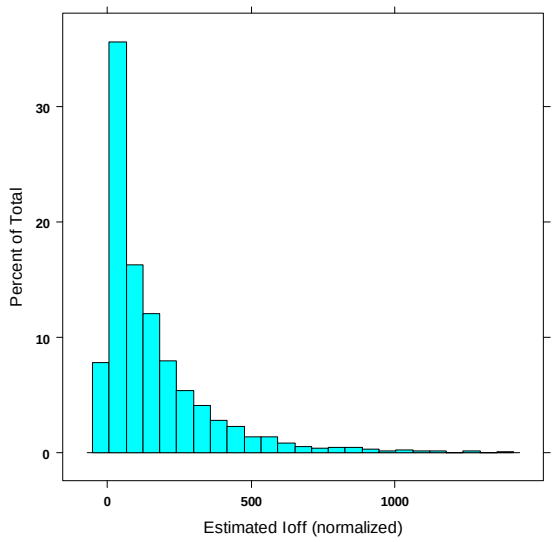
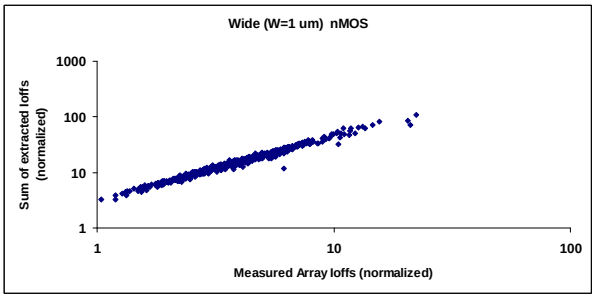



Figure 3: Distribution of extracted $I_{offs}$ from device arrays measured on scribe line during volume production.



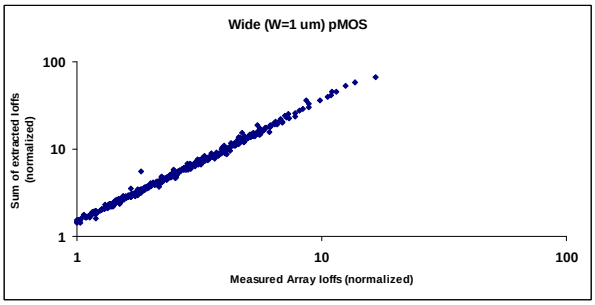
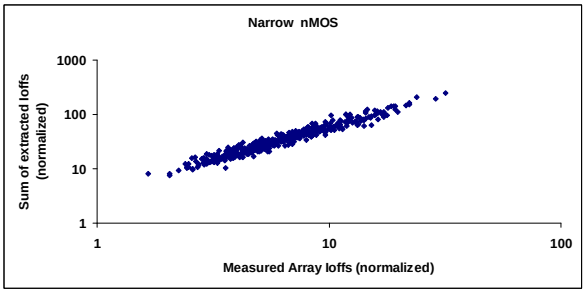




Figure 4: Correlation between sum of extracted $I_{offs}$ of 32 transistor in an array with measured array $I_{offs}$ for wide transistors. Both axes are normalized by the same same constant.



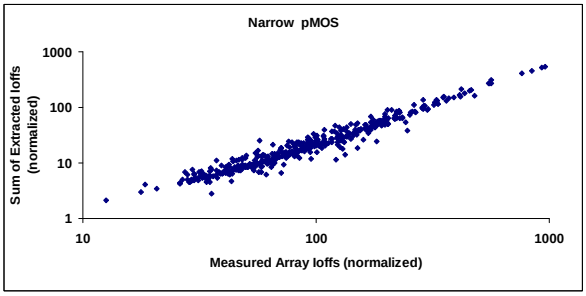




Figure 5: Correlation between sum of extracted $I_{offs}$ of 32 transistor in an array with measured array $I_{offs}$ for narrow transistors. Both axes are normalized by the same constant.

## LEAKAGE CURRENT CALIBRATION

The sum of the extracted $I_{offs}$ from the individual transistors in a transistor array has a very good linear regression with array $I_{offs}$ for any given wafer (the multiple points for regression are the different sites on the wafer where the transistor arrays have been measured). However, this regression does not have slope 1.0 and intercept 0.0. It is possible to extract calibration factors from this slope and intercept for each individual transistor to correct for the contribution of unselected transistors.

Let us say that, for a specific array:

$$\sum_{i=1}^{32} Ioffs_i = Inter + \left(Ioffs_{Array} \times Slope\right),$$ where "$i$" refers to the individual transistors.

We correct each transistor $I_{offs}$ using the formula:

$$Corrected\_Ioffs_i = \frac{\left(Ioffs_i - \frac{Inter \times Ioffs_i}{\sum_{i=1}^{32} Ioffs_i}\right)}{Slope}$$

This assumes that the contribution of individual transistors to the intercept of the regression is proportional to the extracted $I_{offs}$. This calibration can be done for each wafer and each array, resulting in

robust $I_{offs}$ estimates for each individual transistor. Figure 6 shows the comparison between measured array $I_{offs}$ and the sum of extracted $I_{offs}$ for each transistor. A very good match is obtained.

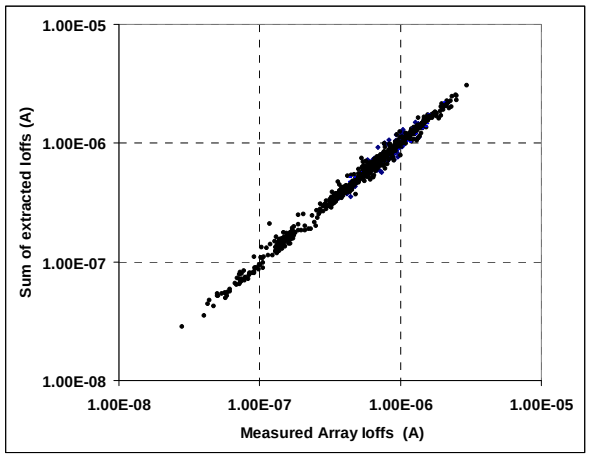



Figure 6: Comparison of sum of extracted $I_{offs}$ after calibration with measured array $I_{offs}$ for multiple device types and dimensions.

## OTHER METHODS FOR LEAKAGE ESTIMATION

A number of test structures have been recently proposed for characterizing leakage variability [4][5][6]. These techniques focus on improving the accuracy of leakage measurements from transistor arrays, where the unselected transistor and selection circuits can contribute to the leakage measured on the shared drain or source pads. The focus of this paper is to make use of low-resolution measurement units for leakage estimation. The motivation for the use of low-resolution measurement units is to enable cost effective fast parallel testing. The leakage estimates obtained by the method described in this paper are suitable for monitoring, control and diagnosis during volume production rather than extremely accurate device characterization and modeling.

The method of extrapolating Id-Vg characteristics in subthreshold to extract $I_{off}$ has also been reported before [7]. One application has been to extract subthershold slope and $I_{offs}$ from Id-Vg curves in the presence of GIDL and parasitic STI transistors. This paper describes another application of the $I_{offs}$ estimation by subthreshold slope extrapolation: cost effective fast parallel test.

## APPLICATIONS

Fast parallel measurement during volume production allows a large sample of measurements to be taken, potentially every die and ever wafer, with minimal test time overhead. This enables rapid diagnosis and improvement of leakage related yield and performance loss.

Figure 7 and Figure 8 show one such application. $I_{offs}$ extraction using the method described in this paper was performed during volume production on transistor array test structures placed in scribe line of a product manufactured in a 65nm technology. This large sample of measurements allowed spatial (across-wafer) analysis of $I_{offs}$ distribution. Figure 7 shows $I_{offs}$ distribution for various spatial zones for one of the nMOS transistors available in this technology. Zone 5 shows increase $I_{offs}$ median and also increased variability in $I_{offs}$. Figure 8 shows product IDDQ for the same wafers. The IDDQ mean and variance is higher in zone 5 compared to other zones. The zonal correlation between $I_{offs}$ and IDDQ helped isolate the causes of IDDQ variability and yield loss to transistor leakage rather than defects.

Figure 9 shows another application, the comparison of $I_{offs}$ distribution of different transistor types offered in the technology. Different transistor types arise due layout and neighborhood differences, Vth, or Vdd differences. In this case, the examination of the distribution of extracted $I_{offs}$ for different transistor types allowed identification of reasons for increase in the tail of the leakage distribution. Figure 9 shows that the distribution of $I_{offs}$ for two transistors types available in this technology. An increase in the tail of the $I_{offs}$ distribution is seen on wafer 5 for both types of transistors. This change in the distribution, which is made apparent by large sample of measurements on multiple transistor types, allows the isolation of the root-cause of leakage related yield loss in this product.

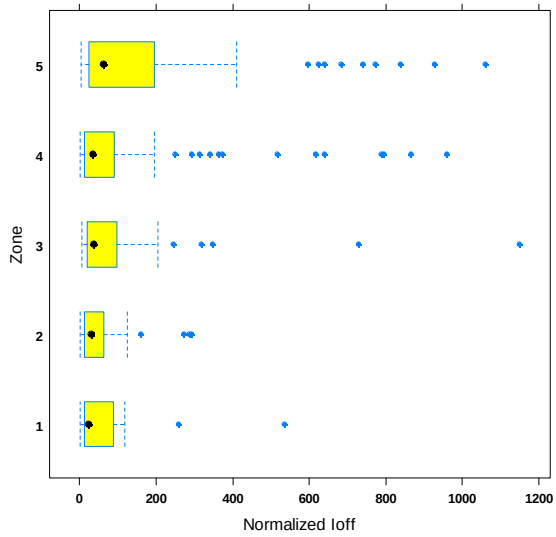



Figure 7: Across wafer spatial analysis of $I_{offs}$ distribution. $I_{offs}$ distribution for each spatial zone is shown as a box-plot.

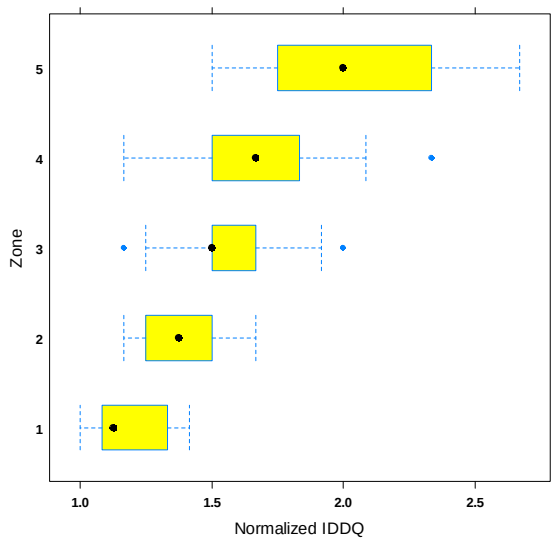



Figure 8: Across wafer spatial analysis of product IDDQ. IDDQ distribution of each spatial zone is shown as a box-plot.

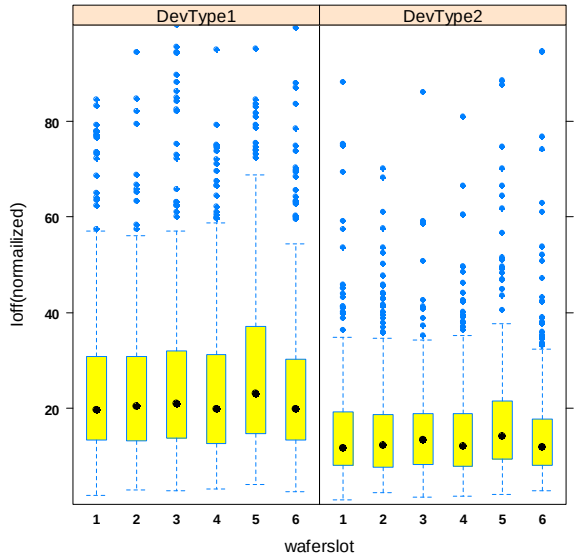



Figure 9: $I_{offs}$ distribution of different device types available in a technology. $I_{offs}$ distribution of each wafer is shown as a box-plot.

## CONCLUSIONS

This paper described a method for using low-cost low-resolution parallel testers to extract off-state subthreshold leakage $I_{offs}$. $I_{offs}$ was extracted from subthreshold slope. The method was applied to both individual and arrayed devices. In the case of individual devices, the extracted $I_{offs}$ was in good agreement with the measured $I_{offs}$ with error less than 10%. In the case of arrayed transistors, for a simple array architecture suitable for scribe line placement utilized in this work, the correlation between the sum of estimated off current and measured array off-current was greater than 0.9 even though estimated $I_{offs}$ did not match array $I_{offs}$. The strong correlation was used to derive calibration factors, which were then used to estimate individual transistor $I_{offs}$ from array test structures. The ability to estimate individual transistor $I_{offs}$ using fast parallel test techniques allows statistical characterization of off-state currents enabling IDDQ estimation and yield improvement.

## REFERENCES

[1] C. Hess, S. Saxena, H. Karbasi, et al. “Device Array Scribe Characterization Vehicle Test Chip for Ultra Fast Product Wafer Variability Monitoring”, IEEE International Conference on Microelectronic Test Structures, pp. 145-149, March 2007

[2] S. Saxena, C. Hess, H. Karbasi, et al., “Variation in Transistor Performance and Leakage in Nanometer-scale technologies”, *IEEE Transactions on Electron Devices*, vol. 55, num. 1, pp. 131-144, 2008.

[3] Y. Taur and T. H. Ning, “Fundamentals of Modern VLSI Devices,” Cambridge University Press,1998

[4] T. Sato, H. Ueyama, N. Nakayama, et al.,, “A MOS transistor array with pico-ampere precision for accurate characterization of leakage current variation”, IEEE Asian Solid-State Circuits Conferemce, pp. 389-392, 2008.

[5] B. Smith, U, Annamalai, A. Arriordaz, et al., “A novel biasing technique for addressable parametric arrays”, IEEE International Conference on Microelectronic Test Structures, pp 166-171, 2008.

[6] K. Agarwal, F. Liu, C. McDowell, et al., “A test structure for characterizing device mismatches”, IEEE Symp. on VLSI Circuits, 2006.

[7] H. Brut and R.M.D.A Velghe, “Contribution to the characterization of the hump effect in MOSFET submicronic technologies”, IEEE International Conference on Microelectronic Test Structures, pp 188-193, 1999.

# Test Structures Utilizing High-Precision Fast Testing For 32nm Yield Enhancement

Muthu Karthikeyan, Louis Medina, and Ernesto Shiling

IBM Systems and Technology Group
2070 Route 52, Hopewell Junction, NY 12533, USA
karthikeyan@us.ibm.com

## ABSTRACT

We describe the development and use of various test structures for 32nm yield enhancement. These DC defect test structures are tested in parallel mode on a functional tester using special V/I and Pico-Amp measurement cards. This new test method provides measurement accuracy as high as ± 10 pA along with up to 9x reduction in test time over conventional parametric testing. The large critical area enables reliable estimation of defect densities by failure mechanism.

***Index Terms*—Yield enhancement, Defect limited yield, Process characterization, Parallel test.**

## INTRODUCTION

In an earlier work [1], we reported on an innovative test methodology using the per-pin Parametric Measurement Unit (PMU) of a functional test platform to test DC defect structures in parallel mode. In this paper, we present our recent work extending the parallel-test methodology even further by utilizing special V/I and Pico-Amp measurement cards that provide measurement accuracy as high as ± 10 pA while still maintaining up to a 9x reduction in test time over conventional parametric testing.

Historically, parametric testers have been used to test defect test structures. A traditional parametric tester allows very high precision voltage or current measurement (of the order of nV/fA), but the measurements are made in serial mode using a matrix switching relay resulting in relatively long test times. In order to reliably measure defect densities down to the ppb level, test structures must have large critical area. However, a single structure such as a serpentine or a via-chain is limited to a certain maximum resistance and, in turn, size, dictated by tester measurement specifications and normal parasitic leakage currents. Therefore, to achieve the desired defect observability for a given process layer, the required area must be covered by many smaller test structures, each individually connected to probe pads. This significantly increases the test time on conventional parametric testers using serial-mode testing. More recently, tester manufacturers have started developing parallel test capability on their parametric testers but the reported test time reduction is only 42% [2]. Nevertheless, it is clear that the best way of reducing cost-of-test, without reducing the vast quantity of data needed for process development, is to improve test efficiency.

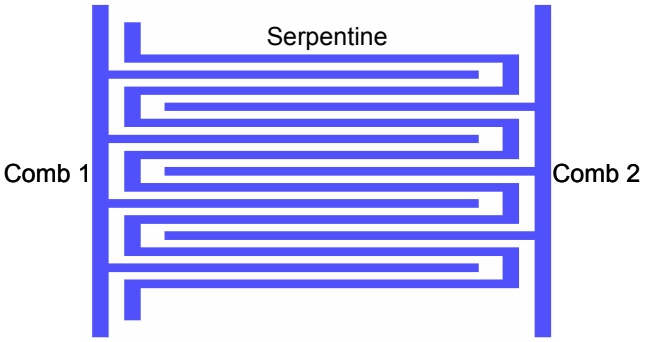
In order to reduce test times without compromising on the quantity or quality of measurement data, we developed a new test methodology to test defect structures in parallel mode using special V/I and Pico-Amp measurement cards integrated into a functional test platform. This setup offers various selectable ranges for current measurement, each with its own guaranteed measurement accuracy and resolution. The use of a digital tester to test yield test structures has been reported previously [3]. In that work, only a binary digit for each DUT representing its pass/fail result based on a preset threshold resistance was collected; the actual resistances of the individual DUTs were not measured. In contrast, we measure the actual current flow through each test structure using the functional test platform, just as in conventional parametric testing, while still achieving up to a 9x reduction in overall test time without any trade-off in the required measurement accuracy. Further test time reduction is possible by optimizing wafer-prober indexing and probe-tip cleaning cycles. The resistance of the DUT is calculated using the measured current and the supply voltage. Based on this calculated resistance value, we are able to classify fails as either hard (catastrophic) or soft (resistive), greatly helping the root cause investigation of yield-loss mechanisms.

## TEST STRUCTURE DESIGN

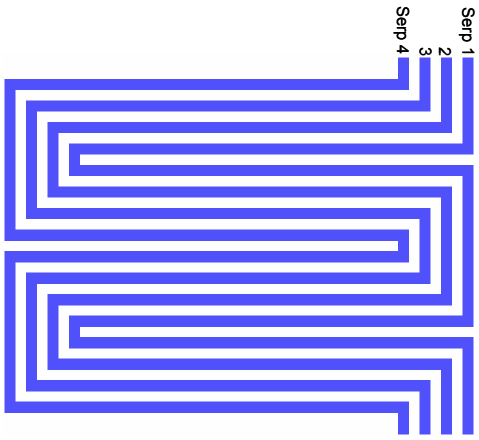
The test structures described in this paper cover all the physical levels on a wafer, namely, Active, Poly, Contact, and up to 11 Metal, and 10 Via levels. They include three types of standard defect test structures – Mazes, Multiple Parallel Serpentines (MPS), and Contact/Via chains.

As shown in Fig. 1(a), a Maze consists of two combs and an intertwined serpentine wire for measuring conductor line opens and shorts defect density, while the MPS shown in Fig. 1(b) consists of 4 parallel wires in a serpentine pattern. By testing each line of the MPS for opens and each pair of

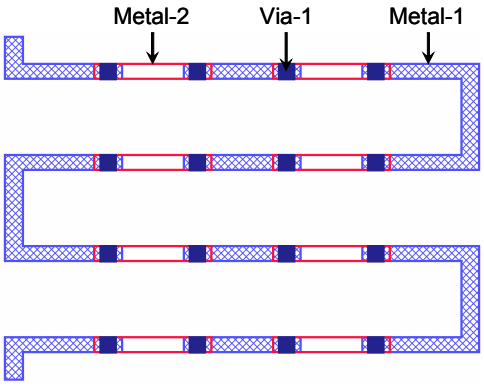
adjacent lines for shorts, the defect size distribution for opens and shorts defects is obtained, similar to [4]. Fig. 1(c) shows the schematic of the Via chain design that is used to test the Via-1 level connecting Metal-1 to Metal-2 above it. Similar chain structures were designed to test the other via levels and also the Contact level that connects Active or Poly to Metal-1.




(a)




(b)




(c)

Fig. 1. Schematic of three types of test structures used in 32nm yield enhancement.
(a) Maze consisting of two combs with an intertwined serpentine wire for measuring line opens and shorts defect density.
(b) MPS consisting of 4 parallel serpentine wires for measuring the defect size distribution for line opens and shorts defects.
(c) Via chain for measuring Via opens defect density (Via-1 in this example).

Many smaller, identical test structures were tiled into a pad frame with probe pads in a 2x50 arrangement as shown in Fig. 2. The reasons for breaking up the test structure area into many smaller test structures, each individually connected to probe pads, rather than designing one large structure are two-fold. First, this arrangement limits the nominal resistance of each test structure to a reasonable level, and, second, reducing the maximum defect area associated with each fail facilitates localization of defects through Physical Failure Analysis (PFA). Electrical testing was done using a 2x50 probe card, and for each touchdown all test structures connected to the 100 pads were measured simultaneously. The Maze, MPS, and Via chain arrays are each 4 to 6 mm$^2$ in area, providing sufficient critical area for accurate characterization of wafer-to-wafer and intra-wafer defect density variation. Tables 1(a) and 1(b) list the total length of wiring per wafer in the Maze and MPS arrays, and Table 1(c) shows the number of vias per wafer in the Via chain array. Currently, these 3 test structure arrays exist on 5 different 32nm masksets, with plans to place them on additional future masksets.

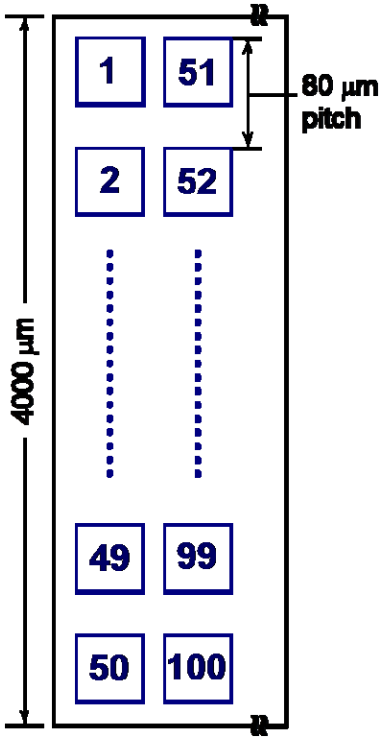



Fig. 2. Standard 2 x 50 pad frame used in the 32nm test structures described in this paper. The width of the test structure array varies from 1 to 1.5 mm.

Table 1. Coverage of the test structures, by level, for the (a) Maze, (b) MPS, and (c) Via chain. This example is for a 32nm test chip with 9 metal levels.

| Planar Conductor level | Design Rule Pitch | Wire length/wafer, per level (m) |
|---|---|---|
| Active | Minimum | 551 |
| Poly | Minimum | 550 |
| Metal-1 through Metal-5 | Minimum | 1,654 |
| Metal-6 and Metal-7 | 2x Min. | 828 |
| Metal-8 and Metal-9 | 8x Min. | 207 |

(a)

| Planar Conductor level | Design Rule Pitch | Wire length/wafer, per level (m) |
|---|---|---|
| Active | Minimum | 944 |
| Poly | Minimum | 960 |
| Metal-1 through Metal-5 | Minimum | 3,227 |
| Metal-6 and Metal-7 | 2x Min. | 1,612 |
| Metal-8 and Metal-9 | 8x Min. | 402 |

(b)

| Via level | Design Rule Size | Number of vias/wafer, per level |
|---|---|---|
| Contact-Active | Minimum | 2.0E+09 |
| Contact-Poly | Minimum | 2.0E+09 |
| Via-1 through Via-4 | Minimum | 4.0E+09 |
| Via-5 | 2x Min. | 1.0E+09 |
| Via-6 | 2x Min. | 5.1E+08 |
| Via-7 | 8x Min. | 1.3E+08 |
| Via-8 | 8x Min. | 1.0E+08 |

(c)

## TEST METHODOLOGY

Historically, parametric testers have been used to test defect test structures. While a traditional parametric tester allows very high precision voltage and current measurements of the order of μV and fA, respectively, the measurements are made serially using a matrix switching relay, resulting in relatively long test times. While some limited improvement in test time is possible by optimizing the measurement algorithms as described in [5], the steep rise in test structure content with every new technology node demands a more radical approach.

In an effort to significantly reduce test time, we developed a novel test methodology using a functional test platform to test defect structures in parallel mode. The hardware solution utilizes three types of resources: Multi-channel V/I Source-Measure modules, custom Pico-Ampere modules, and standard digital modules with a DC Parametric Measurement Unit (PMU) and a test processor embedded in every pin. The Digital module/PMU was the subject of our earlier work [1]. However, in this work, we used the V/I Source-Measure and pA modules exclusively, as they offer a wider current measurement range with better accuracy. Each of these hardware modules uses a custom Wafer Probe Interface (WPI) board to route their resources to the pA module as shown in the block diagram in Fig. 3. This new custom pA measurement module fits into a standard card slot in the functional tester, is water cooled for maximum temperature stability, and is controlled through a Linux USB interface. The pA module controls the individual hardware resources through a relay matrix allowing the user to select any of the 3 hardware resources.

Each of the 100 probe card pins has the full capabilities of the V/I module, the pA module, and the digital module on a per-pin basis. The software solution is fully integrated through the tester operating environment with the use of custom Application Program Interfaces (APIs). Together, the hardware modules and the software provide a seamless measurement capability that includes standard voltage or current measurements, low current measurements in the pA range, and digital/frequency measurements. This solution allows all test structures contained in a 2x50 pad-frame to be tested simultaneously in a single probe touchdown. This setup offers various selectable ranges for current measurement, each with its own resolution and guaranteed accuracy limits. We found the measurement ranges of ± 10 μA (for the V/I card) and ± 200 nA (for the pA card) to be sufficient for testing most of the defect test structures discussed here.

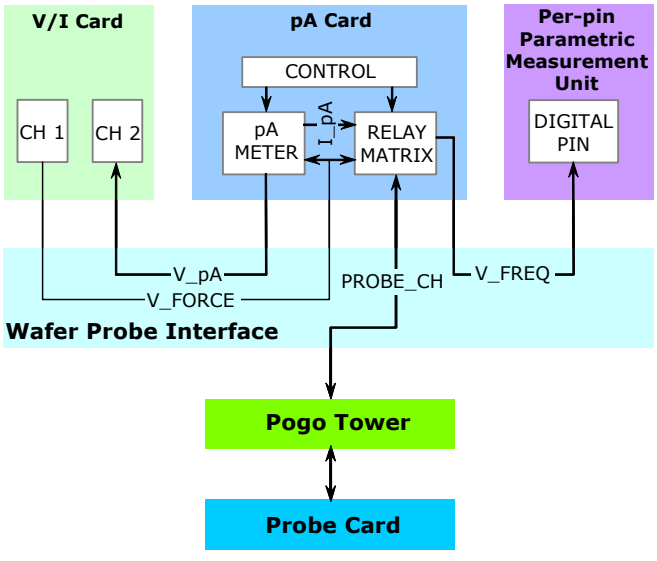
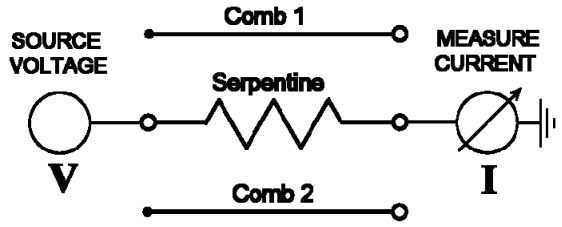
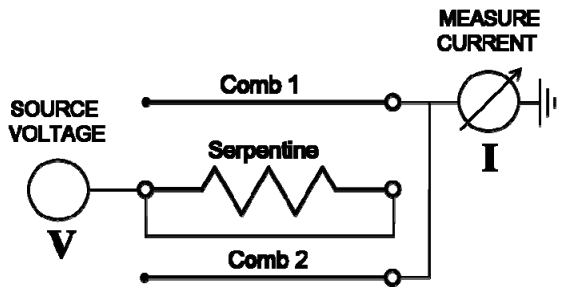



Fig. 3. Block diagram of the various hardware modules in the functional tester with DC measurement capability.

Fig. 4 illustrates the procedure for testing the Maze test structure of Fig. 1(a). For testing line opens, a source voltage of 1.5 V is applied to one end of the serpentine with the other end grounded. The current flow through the serpentine is measured at the ground terminal as shown in Fig. 4(a). The combs are disconnected during opens test. For testing line-to-line shorts, a source voltage of 1.5 V is applied to both ends of the serpentine and the leakage current between the serpentine and the combs is measured at the comb ends which are tied together to ground as shown in Fig. 4(b). The Via chains are tested for opens in the same manner as the Maze, while the algorithm for testing the MPS structure is somewhat more involved.




(a)



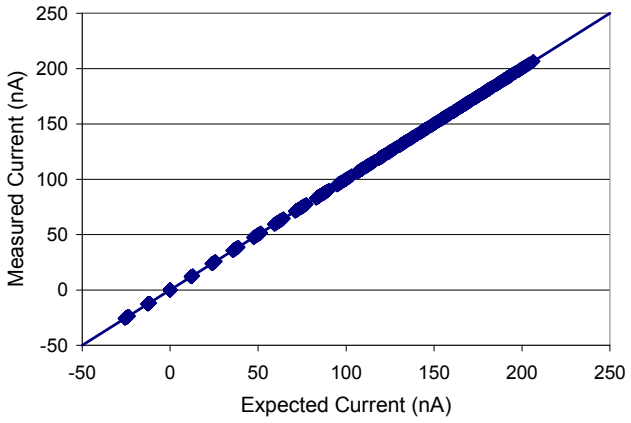

(b)

Fig. 4. Test configuration for testing the Maze structure of Fig. 1(a)
(a) Opens test to detect breaks in the Serpentine wire.
(b) Shorts test to detect bridging between the Serpentine and the Combs.

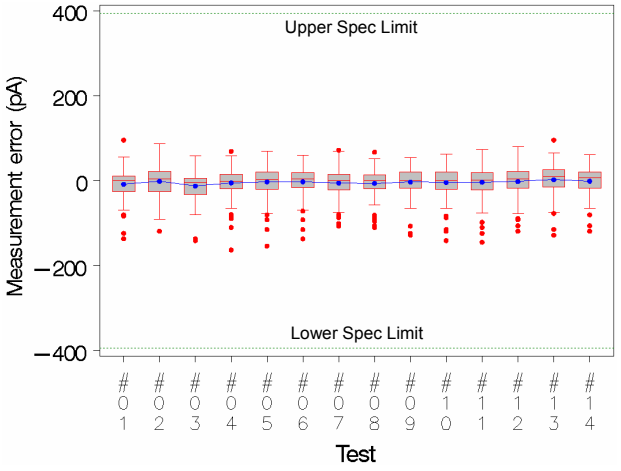
With our parallel-test approach, we achieved up to a 9x test time reduction over conventional parametric testing without any trade-off in the required measurement accuracy. For instance, it took 10.5 minutes to test all the test structure arrays on an entire wafer at the Metal-11 level in parallel mode. However, the corresponding test time for an identical set of structures on a conventional parametric tester is estimated to be 100 minutes. The improvement factor varies somewhat from one test level to another depending on the number of measurements and number of probe touchdowns involved.

## RESULTS AND DISCUSSION

Prior to testing the defect test structures, we conducted a series of experiments to validate the accuracy and repeatability of our parallel-test solution. In these experiments, varying voltages were applied to different standard resistors and the resulting current measured on each of the 100 pins. The measurement set was repeated 14 times over a course of 2 weeks, within one calibration cycle, to gauge long-term measurement repeatability. The expected current is calculated by dividing the supply voltage by the resistance of the standard resistor. Fig. 5(a) shows a scatter plot of the measured current using the pA card, in the measurement range of ± 200 nA, against the expected current. The data in this example was obtained by sweeping the supply voltage from -1.5 V to 8 V in 0.5 V steps across a 40 MΩ standard resistor. The excellent correlation between the actual and expected currents confirms that our parallel-test methodology is sound. Similar experiments were carried out in the other operating ranges. As an example, Fig. 5(b) shows a box plot of the current measurement error (measured current - expected current) for the 1.5 V condition for each one of the 14 tests carried out over 2 weeks. The specification limits for current measurement error are also shown. As seen, the actual measurement error is well within specification.




(a)

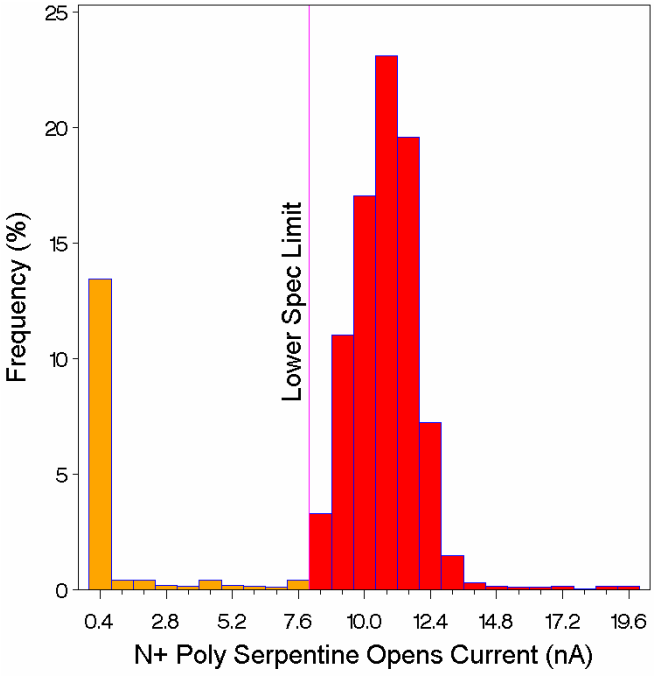



(b)

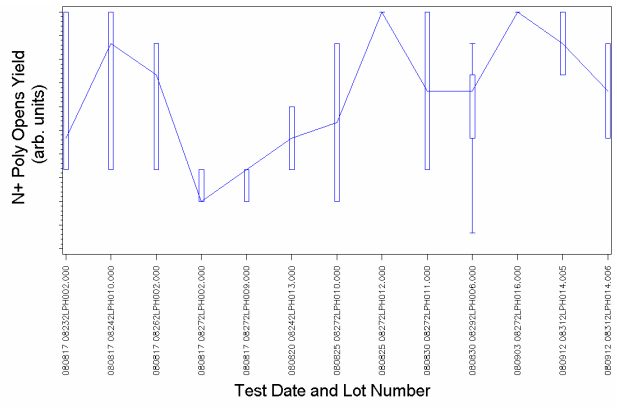
Fig. 5 (a) Measured vs. expected current for the pA measurement hardware in the current measurement range of ± 200 nA.
(b) Current measurement error for 14 sets of measurements carried out over 2 weeks. The specification limits for measurement error are also shown.

As mentioned earlier, the test structures described here currently exist on 5 different 32nm masksets. They are tested at 3 test levels, namely, Metal-1, Metal-4 and top-metal, on every lot. A vast quantity of test data is collected, analyzed, and reported weekly. A full status report on the current 32nm yield detractors and ongoing yield improvement activities is beyond the scope of this paper. However, various examples that highlight both the efficiency of the parallel-test methodology described here and the usefulness of these test structures in 32nm yield learning are provided in this section. Fig. 6(a) shows the distribution of N+ Poly maze opens current measurements for 13 lots, while Fig. 6(b) shows the calculated N+ Poly opens yield, by lot, using an opens current spec of 8 nA for the same dataset. Since we collect and store the raw current measurements, it is possible to distinguish between hard vs. soft opens by varying the spec limit appropriately.

Fig. 7(a) shows the Via-3 chain opens current distribution measured with the V/I card for one lot, while Fig. 7(b) shows a representative wafer map of the opens current of a single Via-3 chain. PFA on one of the failure sites identified an incompletely formed Metal-3 link leading to an un-landed Via-3, as shown in Fig. 7(c).

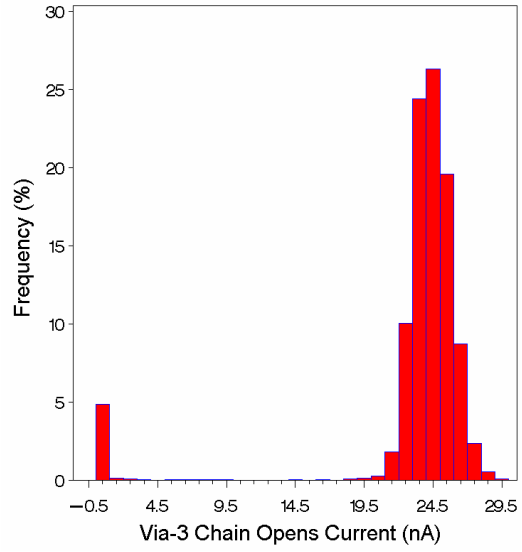



(a)

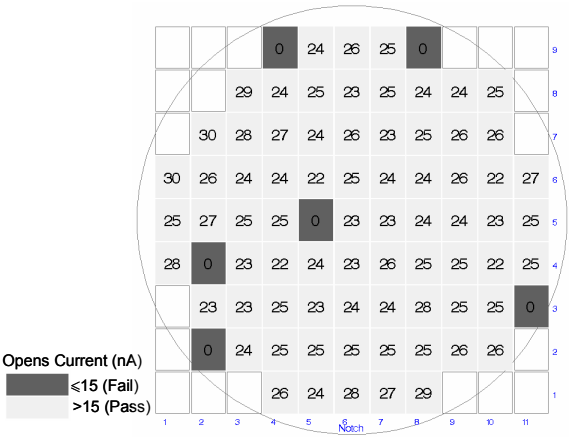



(b)

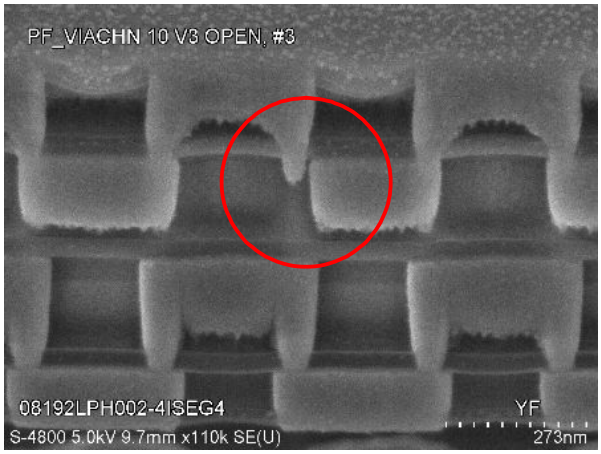
Fig. 6 (a) Histogram of the opens current values of N+ Poly mazes measured on the functional tester using the Pico-Amp measurement card.
(b) N+ Poly opens yield, by lot, calculated from the data shown in Fig. 6(a).




(a)




(b)




(c)

Fig. 7 (a) Histogram of Via-3 chain opens current for one lot measured with the V/I card. The bar on the extreme left represents hard open fails.
(b) Wafermap of Via-3 opens current for a representative wafer in this lot. Physical Failure analysis was conducted on the circled site to identify the root cause of failure.
(c) Cross-sectional SEM image of the localized failure location in the Via-3 chain shows an un-landed Via-3 caused by an incompletely formed Metal-3 link below.

Although the test structures described here were primarily intended to measure overall defect densities, they also offer valuable clues on many systematic yield loss mechanisms. As one example, Fig. 8 shows the difference in Metal-1 through Metal-4 maze opens yield between mazes in the horizontal and vertical orientations for 20 lots. For Metal-4, the opens yield of the vertical maze is significantly lower than that of the horizontal maze, indicating an orientation-dependent Metal-4 patterning issue. As seen below, this yield difference is not present in the other metal levels.

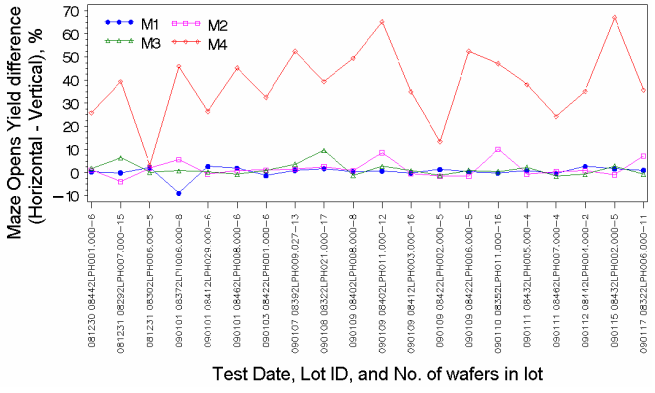



Fig. 8. Maze opens yield difference, by lot, between Horizontal and Vertical mazes for Metal-1 through Metal-4.

Finally, Fig. 9 shows the opens yield trend for the Metal-3 MPS structure of Fig. 1(b), by defect size, for 44 lots. Assuming a perfect circular defect of diameter $s$, the range of defect sizes that cause $n$ adjacent lines to be open can be expressed as a function of MPS line width (= space), $w$. For a 1-line open: $w \le s < 7w$, 2-line open: $3w \le s < 11w$, 3-line open: $5w \le s < 15w$, and lastly for a 4-line open: $s \ge 7w$. The large overlap in these ranges is partly because serpentine wires 1 and 4 turn around on themselves. While smaller defects are expected to be relatively more prevalent, the disproportionately low 1-line opens yield in Fig. 9 indicates a large systematic contributor to overall Metal-3 open yield loss. This phenomenon has also been observed at the other metal levels.

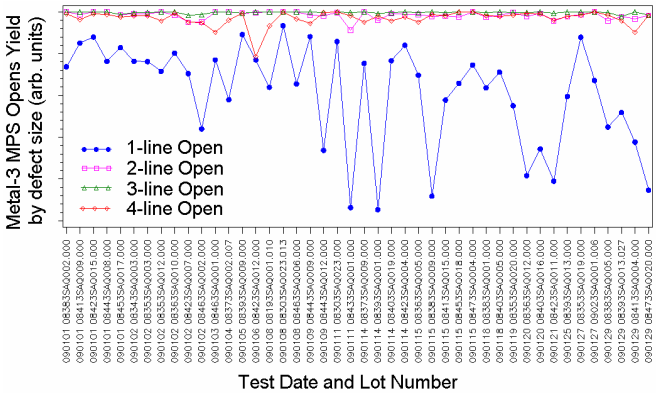



Fig. 9. Metal-3 MPS opens yield by defect size.

## CONCLUSION

A new method for efficiently testing 32nm DC defect test structures on a functional tester has been demonstrated. This parallel-test approach improves test throughput by up to 9x over conventional parametric testing. The test structures help expedite 32nm yield learning through accurate measurement of defect densities by failure mechanism, and characterization of top yield detractors.

## ACKNOWLEDGMENT

We thank Dr. Yi Feng for his assistance with the Physical Failure Analysis.

## REFERENCES

1. M. Karthikeyan, et al., "Short-flow Test Chip Utilizing Fast Testing for Defect Density Monitoring in 45nm," *Proc. 2008 IEEE Int'l Conf. Microelectronic Test Structures*, IEEE Press, 2008, pp. 56-61.
2. J. Kuo, S. Weinzierl, G. Alers, G. Harm, "Parallel test techniques reduce test costs," *Solid State Technology*, November 2004, pp. 32-33.
3. C. Hess, L. Weiland, "Modeling of test structures for efficient online defect monitoring using a digital tester," *Proc. 1994 Int'l Conf. on Microelectronic Test Structures*, IEEE Press, 1994, pp. 108-113.
4. R. Glang, "Defect Size Distribution in VLSI Chips," *IEEE Trans. Semiconductor Manufacturing*, vol. 4, no. 4, Nov. 1991, pp. 265-269.
5. C. Hess, et al., "Test Time Reduction Methods for Yield Test Structures," *Proc. 2003 Int'l Conf. Microelectronic Test Structures*, IEEE Press, 2003, pp. 64-69.

# Efficient Characterization Methodology of Gate-Bulk Leakage and Capacitance for Ultra-Thin Oxide Partially-Depleted (PD) SOI Floating Body CMOS

David Chen, Ryan Lee, Y. C. Liu, Guan Shyan Lin, Mao Chyuan Tang, Meng Fan Wang, C. S. Yeh, S. C. Chien

**Advanced Technology Development Division, United Microelectronics Corporation (UMC)**
No. 3. Li-Hsin Rd. II, Science-Based Industrial Park,
Hsin-Chu City, Taiwan 300, ROC. Fax: 886-3-5785076 Email: david_chen@umc.com

## ABSTRACT

For the first time, an efficient methodology to accurately characterize the gate-bulk leakage current ($I_{gb}$) and gate capacitance ($C_{gg}$) of PD SOI floating body (FB) devices was proposed and demonstrated in 40-nm PD SOI devices with ultra-thin oxide EOT 12A. By applying the RF testing skill for the proposed SOI test patterns, we can eliminate properly the parasitic elements due to the co-existence opposite poly gate type the SOI T-shape body-tied (BT) device and accurately characterize and model the SOI FB $I_{gb}$ and $C_{gg}$ behaviors. Impact on the history effect was analyzed by BSIMSOI4.0 model. History effect analysis with high pulse and low pulse width was shown. Improvement of more than 3% simulation accuracy for history effect was also demonstrated.

## INTRODUCTION

Recently, emerging SOI technology emerges becomes a very promising alternative for the superior power savings characteristics in 40nm and beyond CMOS technologies. One of the key and crucial factors to have a successful design with SOI technology is to characterize and model the FB hysteretic delay variation caused by gate current [1] and capacitance, including $I_{gb}$ and $C_{gg}$. However, in SOI device patterns used for device behavior characterization and modeling, BT with T-shape poly gate devices are used and there exists opposite gate type in the head portion of T-shape poly gate; this will cause different tunneling current mechanism and flat-band voltage. The parasitic elements of $I_{gb}$ and $C_{gg}$ from the head portion of T-shape poly gate [2] cannot be eliminated from the direct AC/DC measurement; and SOI model based upon these device behaviors results in inaccurate simulation for SOI circuit design that includes FB devices.

In this paper, for the first time, an efficient methodology to characterize FB SOI $I_{gb}$ and $C_{gg}$ was discussed and proposed to overcome this impact on device modeling. A set of SOI BT devices having different device dimensions with calibration test patterns were used to characterize $I_{gb}$ and $C_{gg}$ of SOI FB devices simultaneously and accurately for BSIMSOI4.0 model extraction. These test patterns were all with RF test configuration.

## DEVICE FABRICATION

These SOI devices were fabricated on (100) surface with <100> channel direction and a brief process flow was shown in **Fig. 1.** A contact etching stop layer (CESL) film with tensile characteristics for nMOSFET to enhance the carrier mobility was used. The oxide thickness EOT is 12A and a drawn 60nm poly gate length (40nm for TEM) is shown in **Fig. 2.**

## METHODOLOGY

### *A. Test Pattern Design*

**Table I** summarized the device dimensions used in this work. **Fig. 3(a)** shows the basic SOI BT device test pattern for $I_{gb}$ and $C_{gg}$ measurement and **Fig 3(b)** shows the calibration pattern. In **Fig 3(a)** and **Fig. 3(b)**, the head of T-shape poly gate has both N+ and P+ implant types. For BT SOI devices, this kind of structure is correct for characterization and modeling purpose. For FB device, there is no head portion of T-shape poly gate and no opposite implant types in the poly gate. For $I_{gb}$, the Gate tunneling current mechanism and flat-band voltage between body tied and FB devices are much different. **Fig. 4(a), Fig. 4(b)** and **Fig. 4(c)** shows the energy band-diagrams of N+ poly on Pwell, P+ poly on Pwell and P+ poly on P+ Si for the calibration patterns. Charges will be over-estimated for nMOSFET and under-estimated for pMOSFET without proper calibration. Derivation of accurate characterization and modeling for FB devices from body tied device data will need the calibration pattern in **Fig. 3(b)**. Both **Fig. 3(a)** and **Fig. 3(b)** test patterns are all with RF G-S-G test configuration shown in **Fig. 5.**

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

*B. Characterization*

All the accurate $I_{gb}$ and $C_{gg}$ characteristics of FB devices need the measurement from **Fig. 3(a)** and **Fig. 3(b)**. **Fig. 3(b)** is used to eliminate the parasitics. For $I_{gb}$ extraction, it needs the $I_{gb}$ direct measurement from **Fig. 3** only and there is no need to use scattering parameters (S-par). But for $C_{gg}$, it will need the S-par for extraction. In addition to de-embed the parasitic capacitance, we can use the RF C-V methodology [3] to prevent the impact of the high leakage current in the ultra thin oxide device.

## RESULT AND DISCUSSION

**Fig. 6(a)** and **Fig. 6(b)** show the $C_{gg}$ versus gate-bulk voltage $V_{gb}$ plots for calibration pattern, BT and FB devices of nMOSFET and pMOSFET. S-par data from the test patterns with RF-CV test structures were transformed into Z-par to extract $C_{gg}$. In **Fig. 6(a)** and **Fig. 6(b)**, the capacitances of BT device, FB device and calibration pattern are normalized to their corresponding gate poly areas. For nMOSFET, normalized $C_{gg}$ of both BT device and FB device are very close. For pMOSFET, normalized $C_{gg}$ of FB device is greater than that of BT pMOSFET in the inversion region and smaller in the accumulation region. However, for the total capacitances, FB $C_{gg}$ is smaller than the BT $C_{gg}$ for both nMOSFET and pMOSFET. **Fig. 7(a)** shows the $I_{gb}$ vs. $V_{gb}$ of the calibration pattern for nMOSFET and pMOSFET and is used to eliminate the parasitics in the BT device to extract the accurate $I_{gb}$ of FB device. **Fig. 7(b) to Fig. 7(e)** show the BT $I_{gb}$ vs $V_{gb}$ and FB $I_{gb}$ vs $V_{gb}$ for nMOSFET and pMOSFET at different temperatures. Temperature effect plays minor impact on $C_{gg}$ but significant impact on $I_{gb}$ in **Fig. 7(b)** to **Fig. 7(c)**, at the inversion regions of nMOSFET and pMOSFET, the trend is different between FB and BT devices. FB $I_{gb}$ is much smaller than that of BT device after proper calibration. For instance, at nMOSFET (pMOSFET) $V_{gb}$=0.6V(-0.6V) and room temperature, $I_{gb}$ is 2E-12 (1.5E-10) uA/um$^2$ for BT device and about 3E-13 (1E-11) uA/um$^2$ for floating device. Apparently, parasitic $I_{gb}$ existing at the head of T-shape is significant. To analyze the impact on AC performance, we use the delay chain to check the history effect because $I_{gb}$ and $C_{gg}$ strongly affect the history. History effect is caused by the body potential change in a transient condition. The definition of history effect is (Tpd1 – Tpd2) / Tpd2 × 100%; where Tpd1 and Tpd2 are the 1$^{st}$ switch and 2$^{nd}$ switch in the wave shape of gate delay of the inverter. For the history effect impact analysis, we need to firstly generate two BSIM4SOI models: one is based upon the data of BT devices and the other is based upon the data of FB devices. A 10-stage delay chain formed by the inverters is created and shown in **Fig. 8**. **Fig. 9(a)** and **Fig. 9(b)** show the delay versus the pulse width for the BT model and FB model. In **Fig. 9(a)**, high pulse means the falling edge of a low-high-low pulse. In **Fig. 9(b)**, low pulse means the rising edge of a high-low-high pulse. The input pulse width is changed from 100ps to 0.1ps. Parasitic $I_{gb}$ and $C_{gg}$ will cause more than 3% delay difference with applied voltage at 1.0V or above when the input pulse width is 100000 ns. The history effect characterization result of non-calibrated FB $I_{gb}$ and calibrated FB $I_{gb}$ was shown in **Fig. 10**. The difference is more than 3%, which is significant for the accurate SOI chip design. The proposed RF test patterns can be implemented into scribeline of Si wafer with width less than 70um for the statistical model data collection and production monitoring purpose [3].

## SUMMARY

An efficient methodology to accurately characterize and model $I_{gb}$ and $C_{gg}$ for PD SOI ultra-thin oxide FB device was proposed and verified. The importance of calibrated $I_{gb}$ and $C_{gg}$ of FB device was demonstrated. One test pattern set with RF test structure can characterize both DC $I_{gb}$ and AC $C_{gg}$ at the same time. Accurate BSIMSOI4.0 FB model based upon this methodology is achievable for advanced PD SOI chip design. Beyond that, this methodology can be applied in the WAT scribe line (width < 70um) for routine monitoring and statistical data collection without any area cost penalty.

## REFERENCES


[1] P. Su, et. al., *Proc. IEEE ISQED*, p999, 2002.
[2] J. Schmitz, et. al., *IEEE Trans. Semiconductor Manuf.*, vol. 17, no. 2, May 2004, pp150 –154.
[3] D. Chen, et. al., *Proc. ICMTS*, pp243-247, 2007.

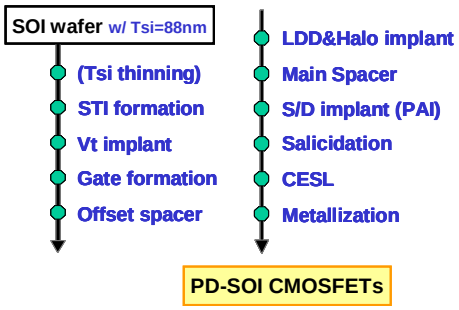




**Fig. 1** Brief process flow of forming PD-SOI CMOSFETs.

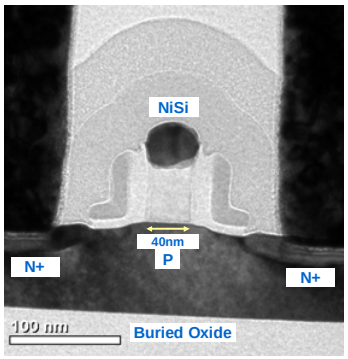
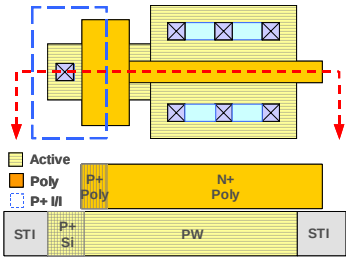
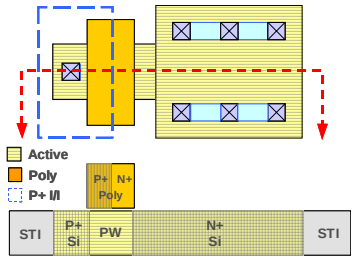



**Fig. 2** TEM picture of PD-SOI device.




**(a)**




**(b)**

**Fig. 3(a)** Basic SOI body-tied (BT) nMOSFET device test pattern for measurements of $I_{gb}$ and $C_{gg}$, and **(b)** the calibration pattern.

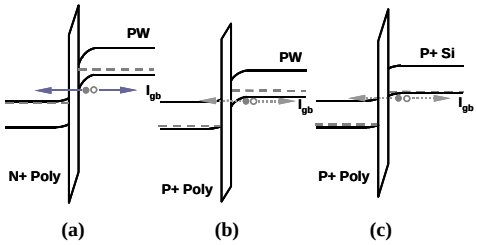



**Fig. 4 (a), (b), and (c)** Energy- band diagrams representing $I_{gb}$ components of the calibration pattern shown in **Fig. 3(b)**.

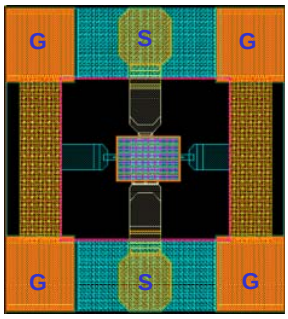
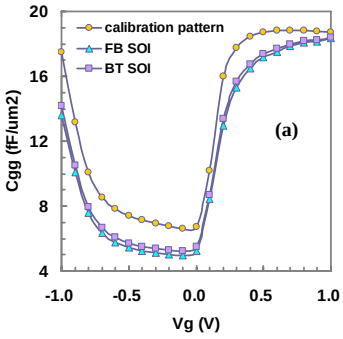



**Fig. 5** Test pattern with GSG pad configuration.



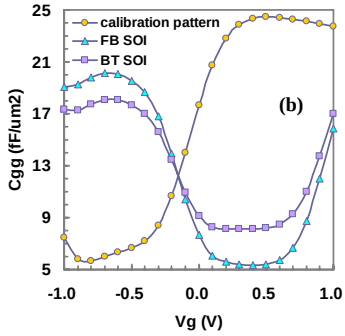
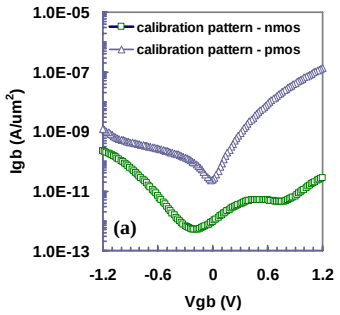
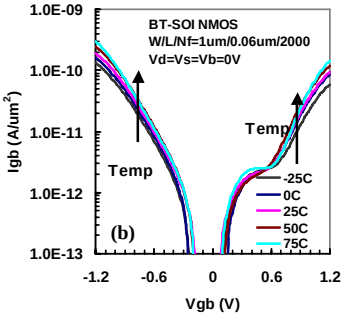



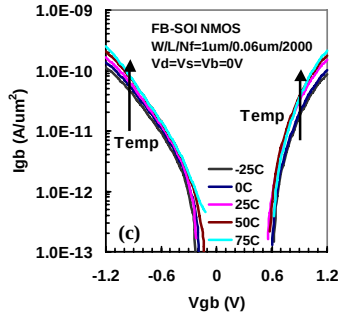

**Fig. 6** $C_{gg}$ versus $V_{gb}$ for calibration pattern, BT and calibrated FB SOI devices for **(a)** nMOSFET and (b) pMOSFET.

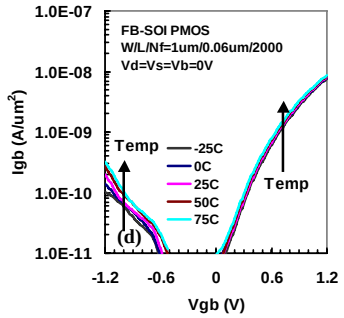
**Table I.** Device W/L/Nf for this work.

| DUT Pattern | W(um) | L(um) | NF |
|---|---|---|---|
| | 2.5 | 0.25 | 1200 |
| | 2.5 | 0.18 | 1200 |
| | 2.5 | 0.13 | 1200 |
| | 2.5 | 0.09 | 1200 |
| | 2.5 | 0.07 | 1200 |
| | 2.5 | 0.06 | 1200 |



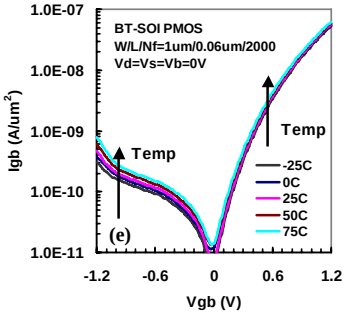




**Fig. 7(a)** $I_{gb}$ vs. $V_{gb}$ of the calibration pattern; **(b)-(e)** BT Igb and calibrated FB $I_{gb}$ vs $V_{gb}$ at various temperatures.

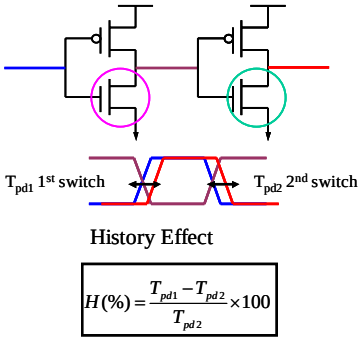



**Fig. 8** Inverter stage of 10-stage delay chain for history effect simulation and history effect definition.

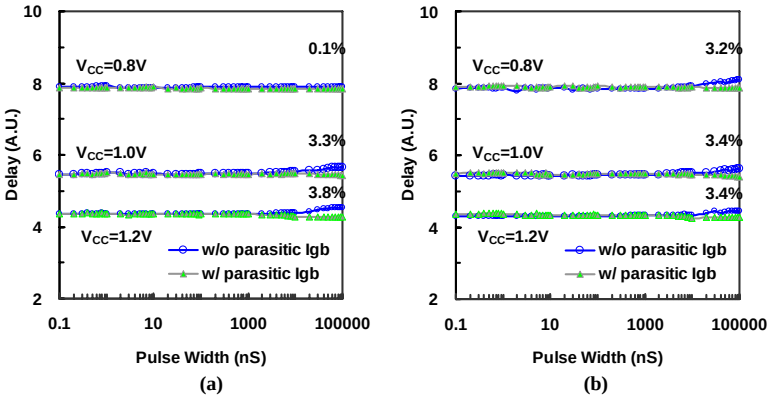



**Fig. 9** Delay versus **(a)** high pulse and **(b)** low pulse width for BT model and FB model.

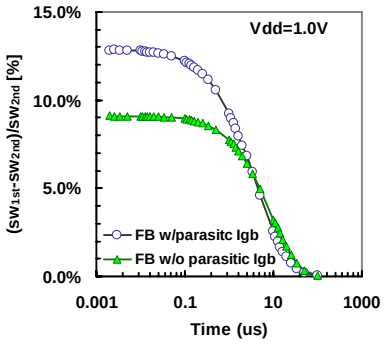



**Fig.10** History effect characterization result with non-calibrated and calibrated FB $I_{gb}$.

# Parameter extraction for the PSP MOSFET model by the combination of genetic and Levenberg-Marquardt algorithms

Q. Zhou, W. Yao, W. Wu, X. Li, Z. Zhu and G. Gildenblat
Ira A. Fulton School of Engineering, Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287, USA

***Abstract*—Based on the combination of the genetic and Levenberg-Marquardt algorithms, a new method is developed to perform both local and global parameter extraction for the PSP MOSFET model. It has been successfully used to extract parameter sets for a 65-nm technology node. Numerical examples demonstrate its ability to obtain highly accurate model parameter values without excessive computational cost.**

## I. Introduction

The accuracy of circuit simulations requires an accurate transistor model and a carefully extracted model parameter set. The PSP model [1], jointly developed by ASU and NXP, is a surface-potential-based compact MOSFET model selected as a new standard for 65-nm CMOS technology and beyond. It has the advantages of an increased physical content, unified expressions for the drain current and terminal charges in all regions of operation and preserves the symmetry essential for several RF applications [2]. Compact MOSFET models are nonlinear functions of the model parameters. A widely used parameter extraction method is the Levenberg-Marquardt (LM) algorithm [3]. It is a gradient-based search algorithm and is very powerful once the starting point is close to the global minimum. Thus choosing the initial point is very important for the algorithm and usually requires considerable experience. On the other hand, genetic algorithm (GA) [4], [5] mimics the natural selection and evolution process and is a non-gradient algorithm. One does not rely on the details of the compact model when using GA. After a large amount of computing, it generates a sub-optimal result from the whole search space. Compared to the LM algorithm, given enough evolution time, GA is more likely to produce a result near the global minimum, but usually is more computationally expensive.

The combination of the genetic and LM algorithms can produce a global minimum after relatively fast computation. The current work presents application of the combined GA/LM algorithms to the problem of parameter extraction for PSP reflecting the unique nature of the PSP scaling equations.

The application of GA to the parameter extraction problem for advanced compact models can also be found in [7] for BSIM3, [8] for SP [9], and [10] for HiSIM. Apart from reflecting the specific nature of scaling in PSP, this work combines GA with LM and quantitatively demonstrates the need for and the use of some of the most important parameters in PSP.

Recently another automatic parameter extraction technique based on particle swarm optimization (PSO) was successfully applied to the local PSP model [6]. As compared to GA, PSO may have the advantages of simpler code and faster convergence. However, with faster convergence, comes an increasing probability that parameter extraction process will stop at a local minimum. On the other hand, by adjusting the details of PSO (e.g. switching between the "gbest" and "lbest" schemes), one can ensure that this technique passes local minima but with the price of slow converge. A detailed comparison between GA and PSO in the context of PSP parameter extraction is a subject of the ongoing investigation.

## II. Local and global model parameters in PSP

To simplify model parameter extraction in PSP one distinguishes between local and global parameters. Local parameters apply to a single device dimension and are actually used to compute the electrical characteristics of MOS transistors. In contrast, global parameters apply to all relevant device geometries. If the global parameter set is available then local parameters for each device can be obtained from the global parameter set via scaling equations. Further details can be found in [1] and [12] while here we consider an example important for the purpose of the present study.

In PSP the drain current increase caused by the channel length modulation effect (CLM) is introduced

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

through the multiplication factor [12]

$$\begin{aligned}\zeta &= T_1 \cdot \left[\mathbf{ALP} + \frac{\mathbf{ALP1} \cdot q_{im}}{(q_{im} + \alpha_m \phi_t)^2}\right] \\ &+ T_2 \cdot \mathbf{ALP2} \cdot q_{bm} \cdot \left(\frac{\alpha_m \phi_t}{q_{im} + \alpha_m \phi_t}\right)^2 \end{aligned} \quad (1)$$

where **ALP**, **ALP1**, and **ALP2** are local model parameters; $q_{im}$, $q_{bm}$, and $\alpha_m$ are the unit channel area inversion charge, bulk charge, and linearization coefficient of the symmetric linearization method at the surface potential mid-point respectively [13]. Variables $T_1$ and $T_2$ are functions of terminal voltages and surface potential variations across the channel [1], [12].

The three local parameters in (1) are obtained from scaling equations

$$\mathbf{ALP} = \mathbf{ALPL} \cdot r_L^{\mathbf{ALPEXP}} \cdot (1 + \mathbf{ALPW} \cdot r_W) \quad (2)$$

$$\mathbf{ALP1} = \frac{\mathbf{ALP1L1} \cdot r_L^{\mathbf{ALP1LEXP}} (1 + \mathbf{ALP1W} \cdot r_W)}{1 + \mathbf{ALP1L2} \cdot r_L^{\mathbf{ALP1LEXP}+1}} \quad (3)$$

and

$$\mathbf{ALP2} = \frac{\mathbf{ALP2L1} \cdot r_L^{\mathbf{ALP2LEXP}} (1 + \mathbf{ALP2W} \cdot r_W)}{1 + \mathbf{ALP2L2} \cdot r_L^{\mathbf{ALP2LEXP}+1}} \quad (4)$$

where the bold text on the right-hand sides of the equations indicates global parameters; $r_L = L_{EN}/L_E$ and $r_W = W_{EN}/W_E$ with $L_{EN} = W_{EN} = 10\,\mu$m and $L_E$ and $W_E$ are effective channel length and width respectively.

While the local parameters are obtained from minimizing the error function for a specific device, global parameters cover the whole range of the relevant geometries. They are obtained by fitting local parameters initially extracted for each individual device to scaling equations. Since this fitting is not perfect, the actual values of the local parameters generated from the global parameter set are slightly different from the original set of local parameters.

## III. Algorithm

Given a discrete set of experimental data, we wish to find a set of model parameters which will minimize the error between the experimental data and the predictions of the compact model. The error can be defined as a suitable norm, usually the $l^2$-norm of the discrete space.

The genetic algorithms mainly contain the operations of selection, crossover, and mutation. Depending on applications, there are many variants. The one used in the present work is real-parameter elite-preserving genetic algorithm [5].

**Selection**. The individuals with better fitness values are chosen for the next generation. The fitness function (taking $I_d$ as an example) used in this work is

$$f = G^{-1} \quad (5)$$

and the total error function is in the form of

$$G = \sum_{i=1}^{4} c_i g_i \quad (6)$$

where $c_i$ ($i = 1, 2, 3,$ or $4$) are weighting factors and $g_i$ are error functions defined as

$$g_1 = \left[\sum_{V_g, V_b, V_d} (I_{d,lab} - I_{d,model})^2 / I_{d,max}^2\right]^{1/2} \quad (7)$$

and $g_2$, $g_3$, $g_4$ in similar form but on output conductance $G_{ds}$, the logarithm of drain current $\log(I_d)$, and transconductance $G_m$ respectively. The subscripts *lab*, *model*, and *max* indicate measured and simulated data, and the maximum of the measured data.

The probability for an individual to be chosen is then

$$P_j = f_j \left(\sum_{j=1}^{l} f_j\right)^{-1} \quad (8)$$

where $j$ denotes the index of an individual and $l$ is the population size.

**Crossover**. The selected individuals are crossed over according to the blend crossover method [5]:

$$x_{j,new}^{(t+1)} = (1 - \gamma_j) x_{j,1}^{(t)} + \gamma_j x_{j,2}^{(t)} \quad (9)$$

where $\gamma_j = (1 + 2\alpha) u_j - \alpha$, $u_j$ is a random number between 0 and 1, $\alpha$ is a constant and can be set to 0.5. The subscripts *new, 1, 2* mean child, Parent 1, Parent 2 and superscript $t$ denotes the generation number.

**Mutation**. In each generation, there is a small fraction of individuals whose parameters will be changed randomly although as pointed out by Deb [5], some researchers argued that it is not a necessary step for the blend crossover operation.

The pseudo code for the particular version of GA used in this work can be written as follows.

*1. Initialize model parameters.*
*2. Initialize population and calculate the fitness of individuals from (5).*
*3. Find the top $\epsilon\%$ members.*
*4. for $i$ = 1 to generation $r$*
*Choose parents according to (8), perform crossover according to (9), and mutation.*
*Calculate the fitness of individuals according to (5).*
*Find the top $\epsilon\%$ members and replace the worst $\epsilon\%$ with the previous top $\epsilon\%$ members.*
*If criteria are satisfied, stop, end if.*
*end for.*

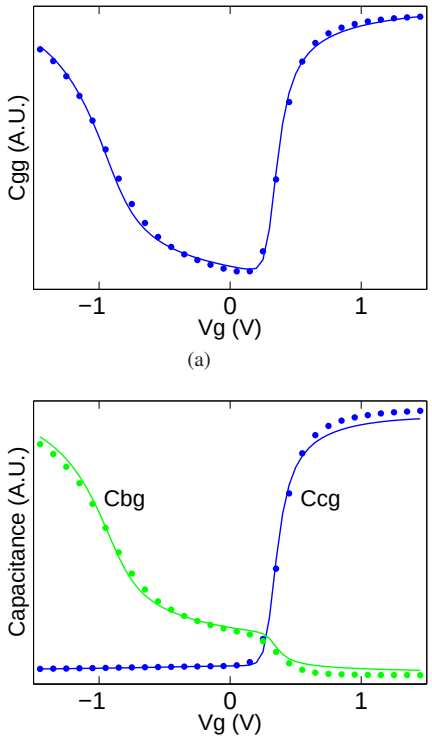



Fig. 1. Transcapacitances (a) $C_{gg}$, (b) $C_{cg}$ and $C_{bg}$ for a device with $W = L = 1\,\mu$m; solid lines are calculated from PSP model while symbols represent the measurements.

LM algorithm is gradient-based and suitable for the situations where the initial point is near the global minimum or where there is only one minimum point in the search space. It can be regarded as a combination of Gauss-Newton and steepest descent methods. For details, see [3].

The search strategy used in this work is that first GA is used to obtain a solution near the global minimum of the error function $G$, the LM algorithm is then applied to further reduce the error function value [8], [11]. The advantage of this combination is that the genetic algorithm leads to a near-optimal solution while the LM algorithm can quickly approach the optimum. If only GA algorithm is used, the computation time will be much longer and if only LM algorithm is used, the solution can be far away from the global optimum.

## IV. Local Parameter Extraction Procedure and Results

Since the implementation time increases exponentially with the number of parameters, it is not practical to extract all the parameters at the same time. Instead, we follow the guideline of the PSP parameter extraction procedure [12]. First the process-related parameters are extracted from C(V) (capacitance-voltage) data. The

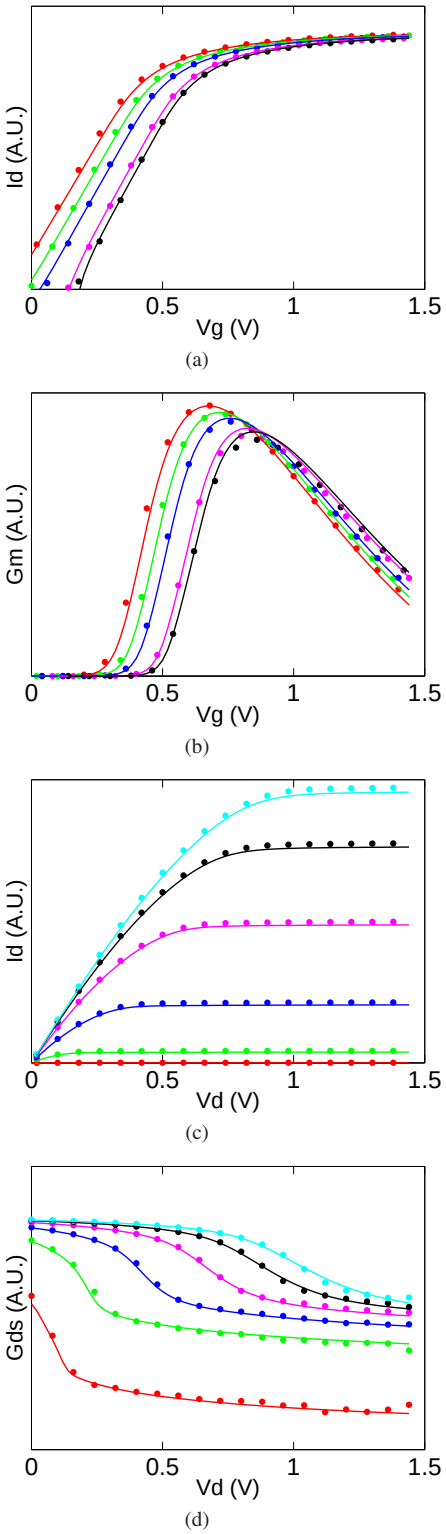
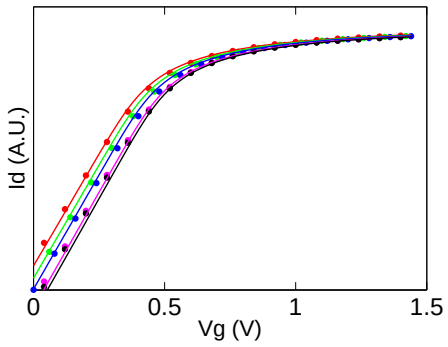
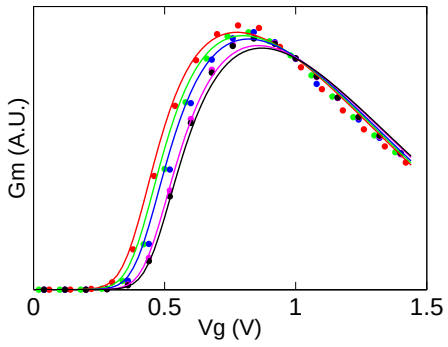


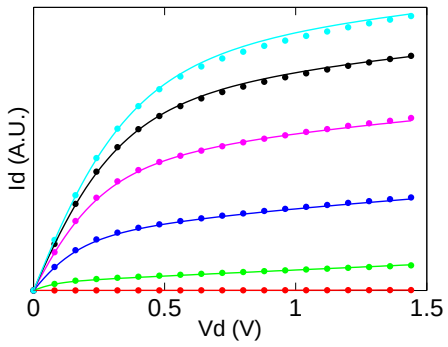

Fig. 2. Drain current, transconductance, and output conductance for a wide-long device with $W = L = 10\,\mu$m; solid lines are obtained from PSP model with global parameter set; symbols represent the measurements. For $I_d$-$V_g$ and $G_m$-$V_g$, the back bias $V_{sb}$ changes from 0 to 0.3, 0.6, 1.2, and 1.4 V (the same for Figs. 3-5). For $I_d$-$V_d$ and $G_{ds}$-$V_d$, $V_g$ changes from 0.3 to 0.6, 0.9, 1.2, 1.44, and 1.6 V (the same for Fig. 3).



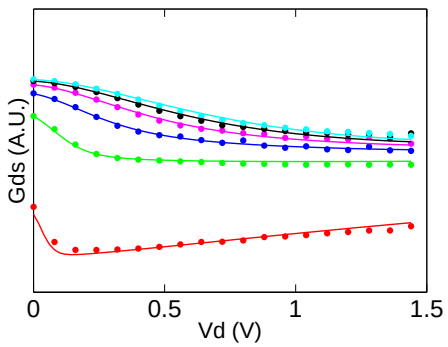

(a)




(b)




(c)




(d)

Fig. 3. Drain current, transconductance, and output conductance for a narrow-short device with $W = 0.12\,\mu$m, $L = 0.06\,\mu$m; solid lines are obtained from PSP model with global parameter set; symbols represent the measurements.

TABLE I
PARAMETERS EXTRACTED AT EACH STEP

| Step | Parameters | | | |
|---|---|---|---|---|
| 1 | VFB | TOX | COX | CGOV |
| | NP | NGOV | | |
| 2 | NEFF | BETN | THEMU | MUE |
| | DPHIB | CT | CS | FETA |
| | RS | XCOR | RSB | RSG |
| 3 | THESAT | CF | AX | VP |
| | ALP | ALP1 | ALP2 | |

mobility related parameters are then extracted from $I_d$-$V_g$ (drain current versus gate voltage) data. Channel length modulation and velocity saturation parameters are from $I_d$-$V_d$ (drain current versus drain voltage) data. Finally extrinsic parameters are extracted from gate current, GIDL, etc. Table I lists the most important model parameters extracted in each step.

The 25 DUTs (Device Under Test) over which we perform the extractions have the geometry range from 10 $\mu$m to 60 nm in length and from 10 $\mu$m to 120 nm in width. The C(V)-related parameters are extracted from two DUTs and once obtained, the parameters are kept intact for all the DUTs.

Fig. 1 shows the capacitance fits between measured (symbols) and simulated (lines) data for a device with $W = L = 1\,\mu$m. Panels (a)-(c) represent transcapacitances

$$C_{gg} = \frac{\partial Q_g}{\partial V_g} \tag{10}$$

$$C_{cg} = -\frac{\partial Q_s}{\partial V_g} - \frac{\partial Q_d}{\partial V_g} \tag{11}$$

and

$$C_{bg} = -\frac{\partial Q_b}{\partial V_g} \tag{12}$$

respectively. Note that $C_{gg} = C_{cg} + C_{bg}$. The PSP model is seen to fit closely with the measured C(V) data.

After C(V)-related parameters are obtained, other local parameters of each DUT are then extracted. Each $I_d$-$V_g$ (or $I_d$-$V_d$) measurement data contain about 440 points. For the GA, the population size is set to around 400 and the number of generations is 500. The search range of each parameter is from the PSP manual [12] but can be modified from experience or tests. A LM algorithm is used to fine-tune the parameters after GA. For a typical DUT, it takes less than 0.5 hours to extract local parameter set on a Dell computer with an Intel 3.73GHz-Xeon CPU.

Comparing the combination of GA and LM algorithms to the purely LM algorithm, we see that it is important to have a good initial guess for LM to obtain a good fit. Table II shows the relative RMS (root-mean-square)

TABLE II
RELATIVE RMS ERRORS (%) FOR DIFFERENT APPROACHES.

| Characteristic | Wide-long device | | Narrow-short device | |
|---|---|---|---|---|
| | GA+LM | LM only | GA+LM | LM only |
| $I_d$–$V_g$ | 2.4 | 17 | 14 | 32 |
| $G_m$–$V_g$ | 2.3 | 15 | 11 | 33 |
| $I_d$–$V_d$ | 0.63 | 2.2 | 1.7 | 4.3 |
| $G_{ds}$–$V_d$ | 7.5 | 8.3 | 11 | 12 |

fitting errors, which are defined as

$$Err = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y_i - s(x_i)}{y_i}\right]^2} \qquad (13)$$

where $m$ is the number of points, $(x_i, y_i)$ are measured data and $s(x_i)$ are simulated data. Since this error function is sensitive to the measurement and numerical noise when the current is very small, the noise-sensitive region is not used in calculation. The results presented in Table II indicate that combined GA/LM algorithm is significantly more accurate than LM algorithm alone. On the other hand, the combined algorithm is about 10-12 times faster than GA alone.

## V. GLOBAL PARAMETER EXTRACTION PROCEDURE AND RESULTS

The PSP local model is designed to fit the electrical behavior of a specific MOSFET device. With geometry scaling included, the PSP global model is aimed at fitting MOSFET transistors within a range of length and width.

Theoretically, GA can be applied to measured data directly to obtain global parameter set. However, this approach is computationally expensive and is not suitable for compact model parameter extraction.

A more efficient approach starts with the local model parameters for different DUTs. A GA or LM procedure is then used to translate the local parameters to the global ones with parameter ranges carefully selected. For complicated scaling equations GA is preferred to LM algorithm since it is not clear where the minimum is located and LM has a very good chance to stop at a local minimum. On the other hand, for simple scaling rules, LM algorithm works well and is computationally efficient. After the translation, LM algorithm is used for the fine tuning. At this point, it is important to cover all the length and width dimensions simultaneously. This is feasible since the ranges of model parameters are already narrowed and the probability of LM converging to a local minimum is reduced. A few iterations of local and global extractions may be needed to obtain good fit.

The fitting of PSP model with thus obtained global parameter set is shown in Figs. 2 and 3. Fig. 2 displays the drain current and conductance fit between measured (symbols) and simulated (lines) data for the wide-long device with $W = L = 10\,\mu$m. Fig. 3 displays the fit for the narrow-short device with $W = 0.12\,\mu$m, $L = 0.06\,\mu$m. The fitting results for other 23 DUTs are similar to those shown in Figs. 2 and 3 and indicate good agreement with the measured data.

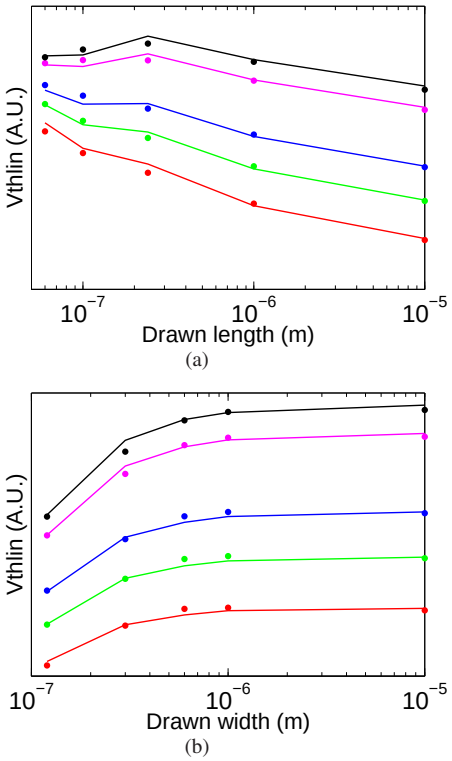



Fig. 4. Linear threshold voltage variations along (a) length and (b) width direction. Lines are calculated from PSP model while symbols represent the measurements.

TABLE III
RELATIVE RMS ERRORS (%) ON $I_d - V_d$ WITH **ALP2** OFF AND ON FOR 3 DIFFERENT DUTS.

| Setting | W=10 $\mu$m L=10 $\mu$m | W=10 $\mu$m L=240 nm | W=10 $\mu$m L=60 nm |
|---|---|---|---|
| Fit with **ALP2**=0 | 1.6 | 2.9 | 2.5 |
| Fit including **ALP2** | 0.63 | 1.7 | 1.8 |

Once the global PSP model parameter set is available, it becomes possible to accurately reproduce the threshold voltage variation with both length and width (Figs. 4 and 5) of the devices. The "linear threshold voltage" plotted in Fig. 4 is defined as the gate bias corresponding to $I_d = 0.3\ \mu\text{A} \times W/L$ and $V_{ds} = 50$ mV. The "saturation threshold voltage" shown in Fig. 5 corresponds to the same current level but with $V_{ds} = 1.2$ V. The narrow-width, short-channel and reverse-short-channel effects are clearly seen in Figs. 4 and 5 and are accurately reproduced by PSP with automatically generated parameter set.

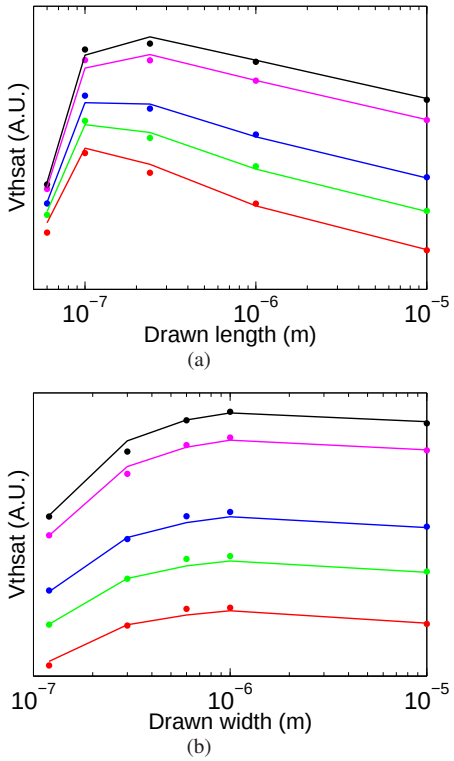



Fig. 5. Saturation threshold voltage variations along (a) length and (b) width direction. Lines are calculated from PSP model while symbols represent the measurements.

## VI. Quantitative Study of the Importance of CLM Parameters

Apart from its application to parameter extraction, the combined GA and LM algorithm is useful in quantitative investigation of the role of (and the need for) specific model parameters. In particular, in this section we investigate the improvement in the fit of test data brought about by inclusion of CLM parameter **ALP2** appearing in (1).

Table III shows the results of the local model fit to the measured data performed with **ALP2**=0 and with **ALP2** included in the optimization process. Physically, **ALP2** represents CLM enhancement factor in the subthreshold region. It becomes clear that the introduction of this factor in compact model significantly reduces the error and that the effect is more pronounced for long-channel device. Note that the relative errors presented in Table III correspond to the optimal choices of all the relevant local parameters which have different values for **ALP2**=0 and optimal **ALP2**.

The quantification of the parameter influence on the model fit to the measured data can be also investigated using the GA algorithm alone (see [8] for the study of coulomb scatting effect). The use of LM algorithm for this purpose is not recommended since convergence to error minimum can significantly skew the results.

## VII. Conclusion

A nonlinear least-square optimization approach with the combination of genetic and Levenberg-Marquardt algorithms has been developed for parameter extraction of the PSP compact MOSFET model. It has the advantages of being flexible, accurate, and automatic and is suitable for the extraction of both local and global parameter sets. The execution time is shorter than that by the GA alone due to the final-stage application of the LM algorithm. The automatic model fits agree well with the experimental data from a 65-nm CMOS technology node. The same technique is applicable to the quantitative investigation of the impact of individual model parameters on the model fit.

## References

[1] 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 Transaction on Electron Devices*, vol. 53, no. 9, 2006, pp. 1979-1993.

[2] X. Li, W. Wu, A. Jha, G. Gildenblat, R. van Langevelde, G.D.J. Smit, A.J. Scholten, D.B.M. Klaassen, C.C. McAndrew, J. Watts, M. Olsen, G. Coram, S. Chaudhry and J. Victory, "Benchmarking the PSP compact model for MOS transistors," *Proc. IEEE International Conference on Microelectronic Test Structures*, March 2007, pp. 259-264.

[3] D.W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathemaitcs*, vol.11, no. 2, 1963, pp. 431-441.

[4] D. Goldberg (1989), *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley.

[5] K. Deb (2001), *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, New York.

[6] R. Thakker, N. Gandhi, M. Patil and K. Anil, "Parameter extraction for PSP MOSFET model using particle swarm optimization," *International Workshop on Physics of Semiconductor Devices, 2007*, pp. 130-133.

[7] J. Watts, C. Bittner, D. Heaberlin and J. Hoffman,"Extraction of compact model parameters for ULSI MOSFETs using a genetic algorithm," *Technical Proceedings of the 1999 International Conference on Modeling and Simulation of Microsystems*, pp. 176-179.

[8] X. Cai, H. Wang, X. Gu, G. Gildenblat and P. Bendix, "Application of the genetic algorithm to compact MOSFET model development and parameter extraction," *Nanotech 2003*, vol. 2, 2003, pp. 314-317.

[9] G. Gildenblat, H. Wang, T.-L. Chen, X. Gu and X. Cai, "SP: An advanced surface-potential-based compact MOSFET model," *IEEE J. Solid-State Circuits*, Vol. 39, Sept 2004, pp. 1394-1406.

[10] M. Murakawa, M. Miura-Mattausch and T. Higuchi, "Towards automatic parameter extraction for surface-potential-based MOSFET models with the genetic algorithm," *Proceedings of the 2005 Conference on Asia South Pacific Design Automation, Shanghai, China, January 18-21*, pp. 204-207.

[11] G. Antoun, M. El-Nozahi, W. Fikry and H. Abbas, "A hybrid genetic algorithm for MOSFET parameter extraction," *IEEE Canadian Conference on Electrical and Computer Engineering, 2003.*, vol. 2, pp. 1111-1114.

[12] X. Li, W. Wu, G. Gildenblat, G.D.J. Smit, A.J. Scholten, D.B.M. Klaassen and R. van Langevelde, *PSP 102.3*, 2008. [Online]. Available: http://pspmodel.asu.edu/downloads /psp102p3_summary.pdf

[13] T.-L. Chen and G. Gildenblat, "Symmetric bulk charge linearisation in charge-sheet MOSFET model," *IEE Electronics Letters*, Vol. 37, No. 12, 7 June, 2001, pp. 791-793.

# Characterization and modeling of mechanical stress in silicon-based devices

A. Spessot(a)*, A. Colombi (a,b), G. P. Carnevale (a), and P. Fantini(a)

(a) Numonyx, R&D – Technology development, via Olivetti 2, 20041 Agrate Brianza, Italy
(b) Dipartimento di Elettronica ed Informazione, Politecnico di Milano – 20133 Milan, Italy
(*) alessio.spessot@numonyx.com

## ABSTRACT

In this paper we show a self-consistent methodology to characterize the stress-induced mobility variation in silicon-based devices. The synergy among different experimental techniques (the application of an external mechanical stress and the measure of the process-induced stress), theoretical calculations (based on the finite elements method and the band structure calculation), and silicon validation (given by particular sets of test structures) is the strength of the characterization tool we propose.

## INTRODUCTION

Strained technology is being promoted as the best way to extend the performance of CMOS transistors. As a consequence, the control of the process-induced mechanical stress has recently attracted much attention. Furthermore, also the stress induced degradation in a stacked multichip or the evaluation of the package-related induced stress have to be taken into account [1]. In this paper we show a self-consistent methodology to characterize the stress-induced mobility variation in silicon-based devices. The synergy among different experimental techniques (the application of an external mechanical stress and the measure of the process-induced stress), theoretical calculations (based on the finite elements method and the band structure calculation), and silicon validation (given by particular sets of test structures) is the strength of the characterization tool we propose .

## EXPERIMENTAL APPARATUS

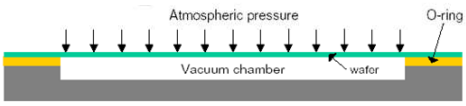



Fig. 1. Sketch of the pressure-controlled kit. Wafer bending can be induced controlling the pressure in the vacuum chamber.

We have developed two different configurations of mechanical stress kits, in order to apply uniaxial and biaxial stress. With two different apparatus, we are able to induce, at wafer level, a compressive or a tensile stress[2,3].
The compressive stress is induced via a pressure-regulated deformation chamber (see Fig. 1). A selected vacuum is induced in the chamber with a pressure controller, and, as a consequence, the atmospheric pressure causes the bending of the wafer.
The tensile stress is applied to the sample with a mechanical kit, able to induce the wafer bending via the four-point-bending technique[4]. Fig. 2 shows the conceptual pattern of the tensile effort experiment. The wafer bending can be controlled fixing his perimeter with a circular holder, and pushing upward his center with a pivot.

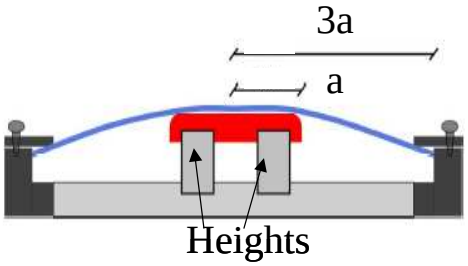



Fig. 2 Sketch of the pivot-controlled kit. The heights of the pivots can be modified, in order to control the wafer bending

## STRUCTURAL DEFORMATION

The induced wafer bending is computed with the elasticity theory, through a mechanical simulation based on the finite element method (FEM). The relation between the resulting strain components $\varepsilon_{ij}$ and the displacement is calculated accounting for the large deformation by the Green-Lagrange strain equation[5] :

$$\varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial R_i}{\partial x_j} + \frac{\partial R_j}{\partial x_i} + \frac{\partial R_k}{\partial x_i} \cdot \frac{\partial R_k}{\partial x_j}\right) \quad (1)$$

where the $R_i$ represents the displacement components along the $x_i$ directions, and the third term takes into account the non-linearity. After the strain calculation, we compute the stress in the material using the constitutive relation[6] $\boldsymbol{\sigma} = \mathbf{D}\boldsymbol{\varepsilon}$ , $\mathbf{D}$ being the 6x6 elasticity matrix and $\boldsymbol{\sigma}$ the stress tensor. To simplify the bending calculation, we introduced a penality contact instead of a step function contact formulation.

So, we are able to calculate the strain state in the *real* device, taking into account the intrinsic anisotropy of the problem (due to the silicon properties) and its non linearity (given by the contact). The resulting applied stress with two configurations is visible in Figs 3 and 4, with the pressure-controlled kit (Fig. 1) and the pivot-controlled kit (Fig. 2).

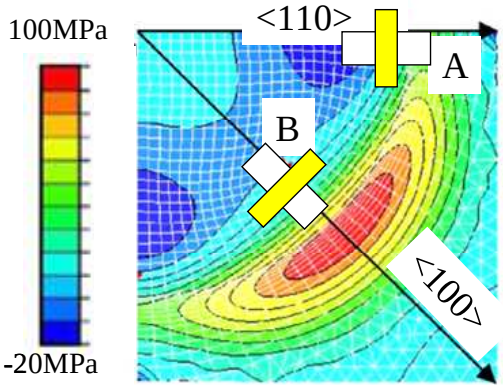



Fig. 3. FEM simulation of the resulting stress obtained with the uniaxial configuration (Fig. 1). MOSFET device are grown on the (001) direction and oriented along the <110> crystallographic axis. They are measured in the maximum stress intensity position (shown as A and B).

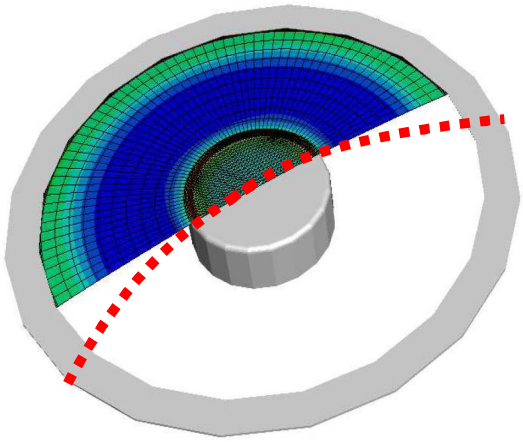


Fig. 4 FEM simulation of the wafer bending obtained with the kit of Fig. 2. For simplicity, only 1/2 of the entire wafer is shown.

## ELECTRONIC SIMULATION

As it concerns the conduction band modification, we adopt the deformation-potential theory defining a rigid shift of its bottom edge produced by a homogeneous strain[7] :

$$\delta E_{nk} = \sum_{j=1}^{6} \Xi_j \cdot u_j \,. \qquad (2)$$

where $\delta E_{nk}$ is the shift of the energy at the band edge point $k$ due to a strain with components $u_j$ referred to the crystallographic axes; the deformation potential terms $\Xi_j$ are calculated according to the deformation potential at the zone bondaries ($\Xi_U$=10.5 eV, $\Xi_D$=1.1eV). The variations in the electron population are derived using the following equation:

$$\frac{\delta n_{nk}}{n_{0nk}} = -\frac{1}{kT} \cdot \delta E_{nk} \,. \qquad (3)$$

where the energy shifts are calculated according to the (2). Fig. 5 reports the so determined energetic shift for the conduction band as a function of stress.

For the valence band, we use the Bir-Pikus theory[8] considering both the degeneracy lift between light holes (*LH*) and heavy holes (*HH*) and the bands dispersion deformation, computed through the following equation:

$$E_{LH}(k) = Ak^2 + a\Delta + \sqrt{E_k + E_{\varepsilon k} + E_{\Delta k}} \qquad (4)$$

$$E_{HH}(k) = Ak^2 + a\Delta - \sqrt{E_k + E_{\varepsilon k} + E_{\Delta k}} \qquad (5)$$

where the term $A$, $a$ and $E_k$ are related to the undeformed bulk silicon, while the terms $E_{\varepsilon k}$ and $E_{\Delta k}$ depend on the strain tensor. The computed band deformation is visible in Fig. 6, where a compressive stress along $W$ is applied. The resulting variations for both $n$ and $p$ channel are according with the piezoelectric theory. Therefore, from the external mechanical bending we can tune the piezoelectric coefficient of the real device, obtaining the low field mobility variations by the following equation:

$$\frac{\Delta\mu_{\parallel}}{\mu_0} = \pi_{\parallel}\sigma_{\parallel} \qquad \frac{\Delta\mu_{\perp}}{\mu_0} = \pi_{\perp}\sigma_{\perp} \qquad \frac{\Delta\mu_{biax}}{\mu_0} = \pi_{biax}\sigma_{biax} \qquad (6)$$

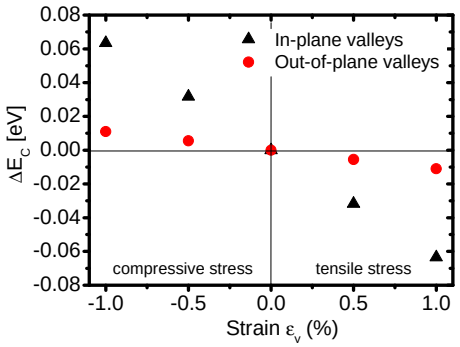



Fig. 5 NMOS: Shift of the conduction band minima, induced by the silicon stress, calculated for the in-plane and the out-of-plane valleys with the deformation potential theory (Eq. 3, see text)

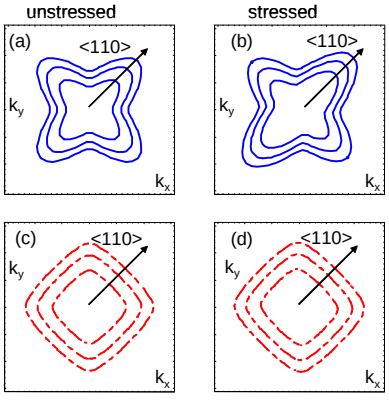



Fig. 6. PMOS: results of the band calculation, showing the stress induced deformation of the band configuration for light holes (LH) and heavy holes (HH) when a compressive stress along W is applied. Eqs (5) and (6) are employed

## CALIBRATION

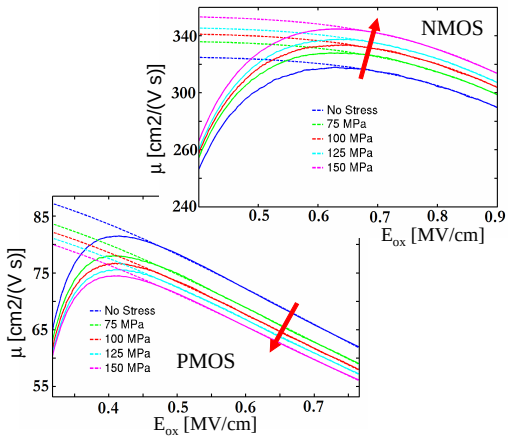



Fig. 7 Universal mobility curves determined by C-V split method, under different amount of stress. The fitting function are able to describe the variations in the mobility. The different effect of the tensile stress on n and p channel is shown by the arrows

Fig. 7 shows the carrier mobility we measured by means of the split-CV method[9], investigating both n-channel and p-channel under different stress directions. Then, we use a fitting function in a selected range of effective vertical field, in order to avoid also the noise source[10] .

We applied an appropriate range of external stress at the sample, measuring the mobility variation at the different vertical electric field. Then, we compared it with the simulation of the mobility variations, obtained by coupling the FEM simulation with the piezoelectric theory, using (6).

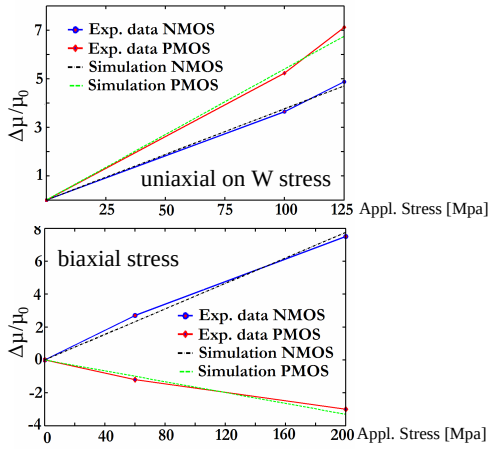



Fig. 8. Comparison between low field mobility variations: experimental data from the mechanical stressed structures and simulations for uniaxial on W and biaxial. Our structural and electronic simulations well capture the experimental trend.

Fig. 8 shows the good agreement among the FEM simulations, band structure calculation and the experimental mobility variation data at various intensity of stress, for different crystallographic stress directions. The calibrated piezoelectric coefficients are given in the following table:

| | $\Pi_{\parallel}$ | $\Pi_{\perp}$ | $\Pi_{biax}$ |
|---|---|---|---|
| *N* | 550 | 490 | 410 |
| *P* | -500 | 710 | 195 |

where the value are given by 10-12 Pa-1, in good agreement with the data reported in literature [4].

## RESULTS

It has been recently suggested that the trend of the $I_{Dsat}$ vs $W$ shows a so called hook-shape[11]. We pointed out that these trends are related to the non-monotonic trend of the gain as a function of $W$, as clearly visible in Fig. 9(a) for both *n*-channel and *p*-channel[12] .

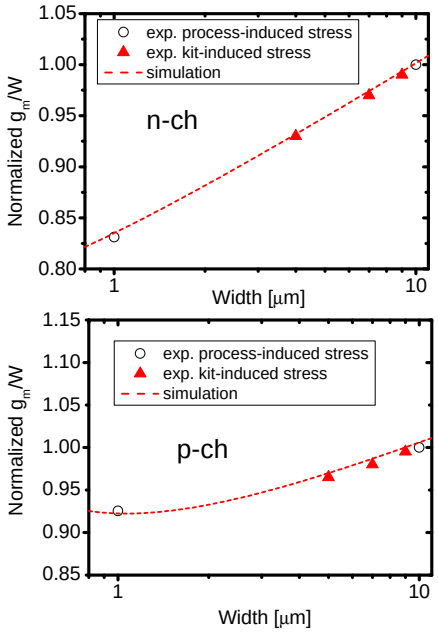



Fig. 9. Trend of $g_m/W$ vs W for *n*-channel (a) and *p*-channel (b). Calculated mobility variation are reported, in comparison with experimental data obtained by the applied stress. The experimental point obtained with the mechanical stress are coherent with the process induced stress.

Fig. 9 also shows the experimental data collected with the application of an external compressive mechanical stress reported in the same $W$ scale. On the basis of the same experimental trend we suggest that the stress effect is the main reason of the observed hook-shape, even if, we are aware, there are other possible causes affecting the gain behavior as a function of $W$ (i.e. the STI induced doping segregation and STI corner rounding). In fact, the initial gain lowering with $W$

scaling up to around 1 μm can be ascribed to the low-field mobility decreasing by the lateral STI-induced stress, while the increasing of the gain trend for $W$<1 μm is related to Δ$W$ effect. The process induced stress mobility reduction can be proved with the process simulation and the altered band structure calculation, (in analogy with the calculation of Figs 5 and 6). Using the presented methodology, we are able to correctly reproduce the strain induced variation in the gain (Fig. 9b). Then, the Δ$W$ extraction can be obtained avoiding any strain effects that can affect the extraction.

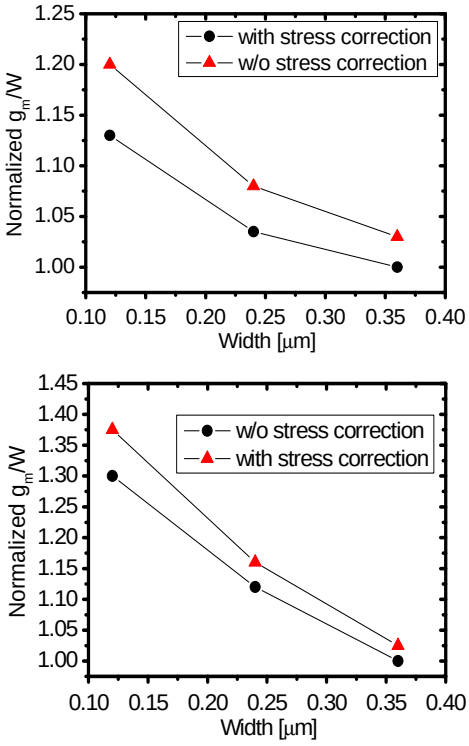



Fig. 10. Normalized gain trend vs W, with (red) and w/o (black) the correction of the stress along W, based on the exposed calibration. The effective ΔW is different by a value of 30% for *n*-channel and 16% of *p*-channel.

Fig. 10 shows the difference between the two approaches: the extraction of the Δ$W$ avoiding the consideration of the strain in $W$ direction causes an error in the Δ$W$ parameter by 30% and 16% for $n$ and $p$ channel, respectively. We proved these results by using a dedicated test structure (Fig. 11). We designed an array of MOSFET, connected each other in order to measure the capacitance. The drain pad of one transistor is singularly connected, allowing the transcurrent measure. Using a series of transistor array with different $W$, we can obtain a Δ$W$ value, which is in agreement with the value extracted from the gain trend of Fig. 9. In the previous literature there are other works involving the application of an external mechanical stress, to characterize the effects of the strain in the carrier mobility. [13-16]. In our knowledge, this is the first paper in which a comprehensive methodology is constructed, and is used to obtain a good modeling of a gain trend in the $W$ direction that completes a previous approach we have attempted [17].

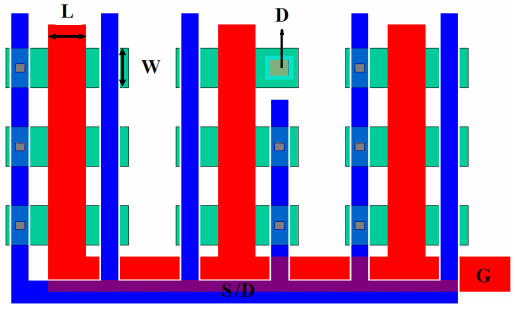



Fig. 11. Test structure used for the validation of the experimental trend. The capacitance can be measured using the Source and Drain (S/D) and Gate (G) pads, which are common for the entire MOSFET array (for a given W, L dimensions). A single device can be measured with the D pad, connecting only one structure.

The measured sample are based on a Flash 65-nm technology node, where some stressor layers are used, in the framework of the strained-silicon process. This self-consistency of our approach suggests that the global stress, induced by our experimental appararus, matches well with the local stress employed in advanced CMOS devices.

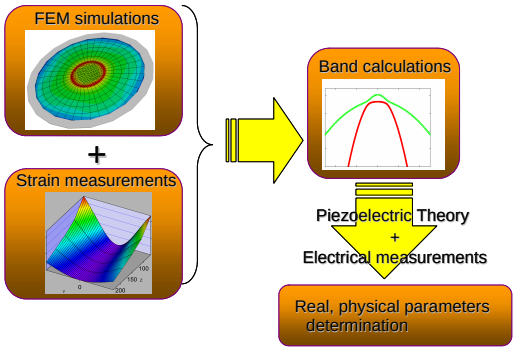



Fig. 12. Summary of the proposed characterization methodology

## CONCLUSIONS

We developed a complete, self-consistent methodology able to characterize the mechanical stress. A schematic view of the method is sketched in Fig. 12. It combines the mechanical and electronic model, promoting a synergy in between that has been validated to study the low field mobility as a function of $W$. In summary, we propose a viable and

comprehensive characterization tool to investigate the performance of strained silicon technology.


### ACKNOWLEDGEMENTS

The authors would like to thank G. Cinà and C. Gerardi (of Numonyx Technology and Device Modelling Group) for fruitful discussions.


### REFERENCES


[1] H. Ali, "Stress-induced Parametric Shift In Plastic Packaged Devices", IEEE Transactions on Components, Packaging B, Vol 20(2) (1997) 458

[2] D. Cordano, G. Carnevale, M. Bocciarelli "Electrical characterization of CMOS transistors subject to externally applied mechanical stress", Comput. Mater. Sci. Vol 43 (2008) Pag 951.

[3] A. Colombi, MD thesis (2008), Politecnico di Milano, Milan

[4] S.E.Thompson, G. Sun, YS Choi, T. Nishida "Uniaxial Process Induced Strained Si: Extending the CMOS Roadmap", IEEE Transactions Electron Devices 53(5) (2006) 1010

[5] J. Lubliner, "Plasticity theory", (1990) Pearson Ed., 2194

[6] S.P. Timoshenko & J.N.Goodier, "Theory of elasticity", McGrawHill ed. (1970)

[7] C. Herring, E. Vogt, "Transport and Deformation-Potential Theory for Many-Valley Semiconductors with Anisotropic Scattering" Phys. Rev. B 101 (1956) Pag. 944

[8] G.E. Pikus & G.L. Bir, Soviet Physics Solid State 1 (1959) 136

[9] Y. Taur T.H. Ning., "Fundamental of Modern VLSI Devices", Cambridge University Press (2002)

[10] K. Chen, C. Hu, P. Fang A. Gupta "Experimental confirmation of an accurate CMOS gate delay model for gate oxide and voltage scaling", IEEE Electron Device Letters 18(6) (1997) 275

[11] P.B.Y. Tan, A.V. Kordesch and O. Sidek, "CMOS Shallow Trench Isolation x-Stress Effect on Channel Width for 130 nm Technology", Solid State and Integrated Circuit Technology (2006), Pag 478

[12] A. Spessot, F. Gattel, P. Fantini, A. Marmiroli, "Modeling of gain in advanced CMOS technologies" Proc. Workshop Compact Modelling (2008), in Press

[13] D. M. Antoniadis, A. Khakifirooz, "MOSFET Performance Scaling: Limitations and future Options", Proc. International Electron Devices Meeting (2008) Pag 253

[14] K. Uchida et al**, "**Experimental study of biaxial and uniaxial strain effects on carrier mobility in bulk and ultrathin-body SOI MOSFETs" Proc International Electron Devices Meeting (2004) Pag. 229

[15] YG Wang et al, "Effects of uniaxial mechanical stress on drive current of 0.13 μm MOSFETs", IEEE Transactions Electron Devices Vol 50 (2003) Pag 529

[16] C. Gallon et al, "Electrical analysis of mechanical stress induced by STI in short MOSFETs using externally applied stress", IEEE Transactions Electron Devices Vol 51 (2004) Pag 1254

[17] Fantini, P.; Giuga, G.; Schippers, S.; Marmiroli, A.; Ferrari, G.;, "Modeling of STI-induced stress phenomena in CMOS 90nm Flash technology" Proc. ESSDERC (2004), Pag. 401

# Improved Parameter Extraction Procedure for PSP-Based MOS Varactor Model

Z. Zhu*, J. Victory†, S. Chaudhry‡, L. Dong‡, Z. Yan‡, J. Zheng‡, W. Wu*, X. Li*, Q. Zhou*, P. Kolev§, C.C. McAndrew¶ and G. Gildenblat*
*Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287
†Sentinel IC Technologies, Irvine, CA 92618
‡Jazz Semiconductor, Newport Beach, CA 92660
§Formerly with RFMD, San Diego, CA 92122
¶Freescale Semiconductor, Tempe, AZ 85284

***Abstract*— We present an improved procedure for extracting parasitic capacitance parameters and gate current parameters for MOSVAR, the industry standard MOS varactor model. Our technique is verified against measured data from three technology nodes (180 nm, 130 nm and 65 nm), and is also used to validate the MOSVAR P-gate/P-well tunneling current sub-model.**

## I. Introduction

For MOS varactor modeling, the Compact Model Council (CMC) has standardized the MOSVAR model model [1], which is based on the PSP MOSFET model model [2]. MOSVAR includes physical models for many phenomena, such as parasitic capacitances and resistances (cf. Fig. 1) and gate tunneling current. These effects are interdependent in real devices; for example, the combined overlap/fringing capacitance and the gate leakage current are correlated through overlap length and the effective channel length and width, which can make parameter extraction challenging [3]. (The overlap and fringing capacitances are lumped together in MOSVAR [1], and below we will use the term "fringing" capacitance to include both overlap and fringing capacitance).

In previous extraction procedures reported for MOSVAR [4] several simplifying restrictions were made, particularly as related to the fringing capacitances. The present work overcomes these restrictions by using a more elaborate analysis of experimental data. Our new method for MOS varactor parameter extraction enables simple determination of effective geometries, fringing capacitances, and the ratio of tunneling currents in overlap and channel regions. This leads to conclusive validation of the MOS varactor tunneling current sub-model [1].

In section II the new extraction methodology for effective geometries and fringing capacitances is introduced. Section III presents an improved procedure for channel and overlap gate current parameter extraction, based on the effective geometry determined in section II. Results from multiple technology nodes are presented. The conclusion follows in Section IV.

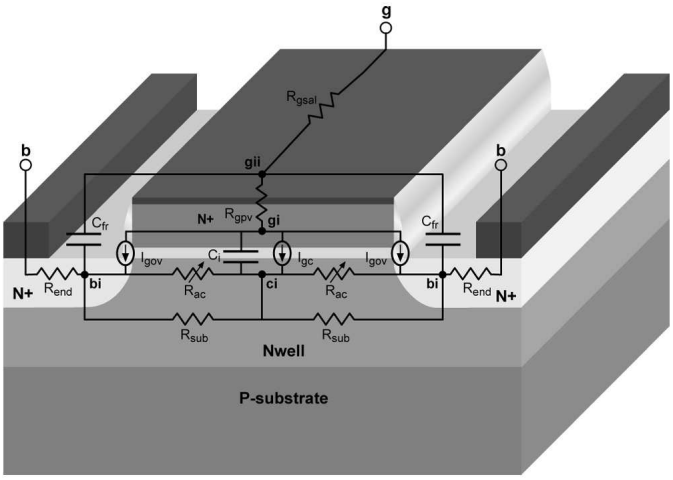



**Fig. 1.** $N^+$-poly/N-well MOS varactor with superimposed MOSVAR model network; after [1].

## II. Fringing Capacitance and Effective Geometry Extraction

The MOS varactor capacitance defined in the MOSVAR model is given by

$$C(V) = C_o(V) \cdot L \cdot W \cdot m + C_{fr} \cdot m \qquad (1)$$

where $C_o$ is the bias-dependent capacitance of the intrinsic part of the device, the length and width are

$$L = L_g + \mathrm{DLQ} \qquad (2)$$

$$W = W_g + \mathrm{DWQ}\ , \qquad (3)$$

and $m$ is the multiplicity factor. The combined fringing and overlap capacitance is given by

$$C_{fr} = 2 \cdot \mathrm{CFRW} \cdot W + 2 \cdot \mathrm{CFRL} \cdot L\ . \qquad (4)$$

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

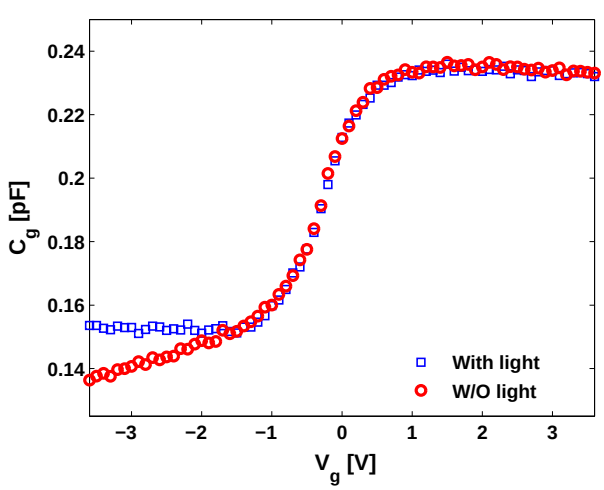



**Fig. 2.** Comparison of measured data between illuminated and dark measurement conditions.

The parameters DWQ and DLQ represent the deviations of the effective channel length $L$ and width $W$ from their "design" or "drawn" values, $L_g$ and $W_g$.

The parameter extraction task is to determine DWQ, DLQ, CFRW and CFRL from test data. In MOSVAR, CFRW encompasses the poly gate overlap of the well contact region ($N^+$ in the case of an N-well) and the fringing capacitance of the poly edge to silicon (as we have noted, a separate term for the overlap component is not included [1]). Previously, the parameters DWQ and DLQ were extracted assuming for simplicity that CFRL $= 0$ [4]. The CFRL term in the MOSVAR model encompasses capacitance associated with the poly gate extension onto STI and the fringing capacitances of the poly edges along the length. Neglecting CFRL presents difficulties in obtaining accurate $C(V)$ and $I_g(V)$ extractions over a wide range of device geometries. In this section we present a new parameter extraction procedure that does not assume that CFRL $= 0$.

Following [4], we consider the minimum and maximum values, $C_{min}$ and $C_{max}$, respectively, of the high frequency capacitance $C$ as determined from S-parameter measurements after de-embedding. To assure identical conditions for measurements and simulations special attention is paid to the frequency behavior of $C$ in the inversion region, which directly influences the value of $C_{min}$. Fig. 2 shows typical $C(V)$ curves measured both with and without illumination. Without illumination the MOS varactor at least partially enters the deep depletion regime, which produces erratic values

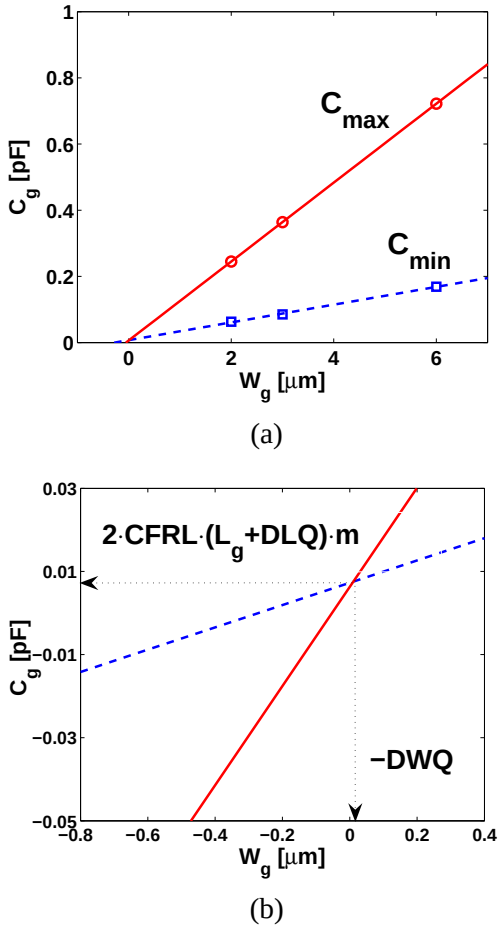
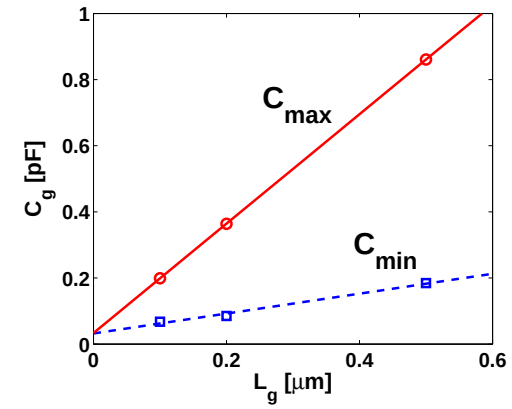



Fig. 3. (a) Measured $C_{min}(W_g)$ and $C_{max}(W_g)$ for a 180nm RFCMOS technology, $L_g = 0.5\ \mu m$, $m = 20$ and $f = 500$ MHz, and (b) Zoom-in of (a) near the intersection point.

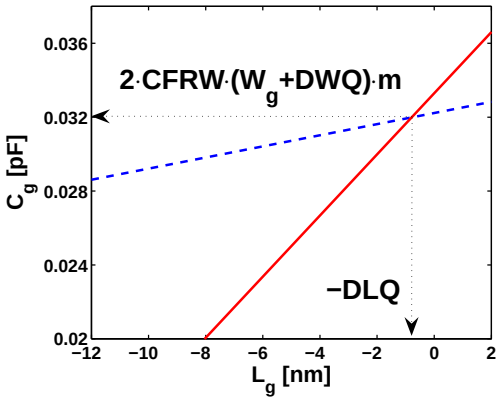
of $C_{min}$ and complicates comparison with simulated results. Under illumination the time required for the inversion layer to form is reduced and a classical high-frequency MOS $C(V)$ curve is experimentally observed. The same $C(V)$ curve is simulated with the MOSVAR v1.0.0 model, which physically includes the frequency dependence of the inversion layer response to the applied voltage [5]. Thus illumination of the sample is required in order to produce consistent measurements of $C_{min}$ for different devices.

Our extraction technique relies on linear regression fitting of $C_{max}$ and $C_{min}$ over geometry. From (1), formally $C_{min}(W_g) = C_{max}(W_g)$ when $W_g = -\text{DWQ}$. Hence at the intercept point of $C_{min}(W_g)$ and $C_{max}(W_g)$ for fixed $L_g$ (as Fig. 3 shows) we have

$$XINT_{wg} = -\text{DWQ}\ , \qquad (5)$$




(a)




(b)

Fig. 4. (a) Measured $C_{min}(L_g)$ and $C_{max}(L_g)$ for a 180nm RFCMOS technology, $W_g = 3\ \mu m$, $m = 20$ and $f = 500$ MHz, and (b) Zoom-in of (a) near the intersection point.

$$YINT_{cwg} = 2 \cdot \text{CFRL}(L_g + \text{DLQ}) \cdot m \ . \qquad (6)$$

Similar linear extrapolation of $C_{min}(L_g)$ and $C_{max}(L_g)$ for fixed $W_g$, as Fig. 4 shows, yields

$$XINT_{lg} = -\text{DLQ} \ , \qquad (7)$$

$$YINT_{clg} = 2 \cdot \text{CFRW} \cdot (W_g + \text{DWQ}) \cdot m \ . \qquad (8)$$

The parameters CFRL and CFRW are computed from (6) and (8), respectively, based on DLQ and DWQ from (7) and (5), respectively.

To verify this procedure, measured and modeled capacitances over bias for different device geometries, which have varying amounts of overlap and fringing capacitance, are shown in Fig. 5 for devices from a 180 nm technology, and in Figs. 6 and 7 for devices from a 130 nm technology. Capacitances over geometry for a 65 nm technology (not shown) are fitted equally as well.

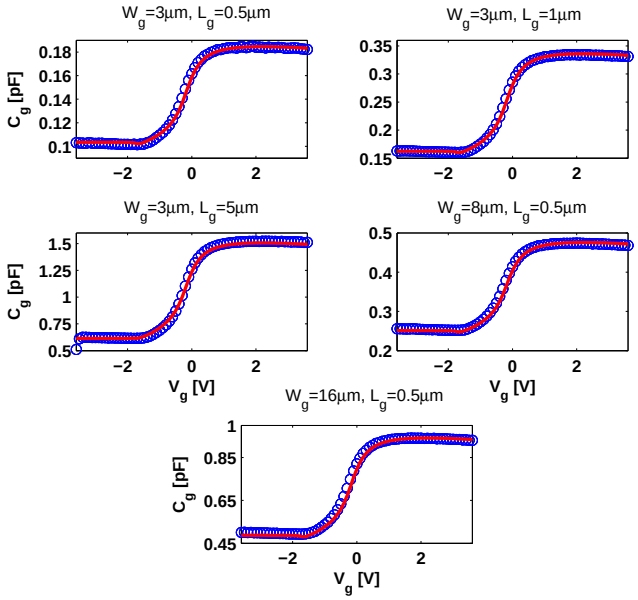



**Fig. 5.** Measured C(V) characteristics at $f = 500$ MHz (open circles) compared with simulation results (solid lines) for different device geometries in a 180 nm tecnology; $m = 20$.

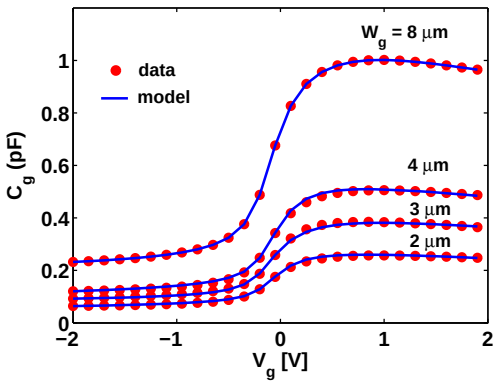



**Fig. 6.** Measured and simulated capacitances at $f = 1$ GHz for various $W_g$ devices in a 130 nm technology; $L_g = 0.5\ \mu m$ and $m = 20$.

## III. Gate Current Parameter Extraction and Model Verification

With DLQ and DWQ determined as described in section II, it becomes possible to find the relative magnitude of the gate tunneling current densities in the channel, $i_{gc}$,

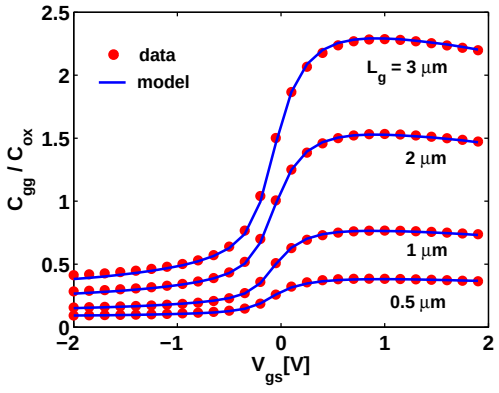



**Fig. 7.** Measured and simulated capacitances at $f = 1$ GHz for various $L_g$ devices in a 130 nm technology; $W_g = 3\ \mu m$ and $m = 20$.

and in the overlap region, $i_{gov}$, contributing to the total gate current

$$I_g = I_{gc} + I_{gov} = i_{gc} \cdot L \cdot W \cdot m + i_{gov} \cdot W \cdot m \ . \quad (9)$$

This involves extracting the parameters IGCHVLW and IGOVHVW entering the expressions [1]

$$i_{gc}(V) = \text{IGCHVLW} \cdot D(V) \cdot F(V) \quad (10)$$

and

$$i_{gov}(V) = 2 \cdot \text{IGOVHVW} \cdot \text{LOV} \cdot D_{ov}(V) \cdot F_{ov}(V) \quad (11)$$

where $D(V)$ and $D_{ov}(V)$ are the tunneling transmission coefficients in channel and overlap regions, respectively, and $F(V)$ and $F_{ov}(V)$ are the associated supply functions [6]. The overlap length LOV $= 10$ nm used in this work was obtained from technology information.

To de-couple the gate current parameters for the channel region from those for the overlap region, $I_g(V)$ data from devices of maximum and minimum drawn channel length, $L_{gmax}$ and $L_{gmin}$, are used. The difference in these currents is

$$\Delta I_g = \text{IGCHVLW} \cdot D(V) \cdot F(V) \cdot (L_{gmax} - L_{gmin}) \cdot W \cdot m \quad (12)$$

which does not depend on the gate current in the overlap region. The parameters IGCHVLW, GCOHVO, GC2HVO and GC3HVO, which model the gate-to-channel region tunneling current, are determined by fitting $\Delta I_g$ defined by (12) (see Fig. 8a).

The total gate current of the minimum length device has the greatest relative contribution from gate current in the overlap region, and so is what should be used

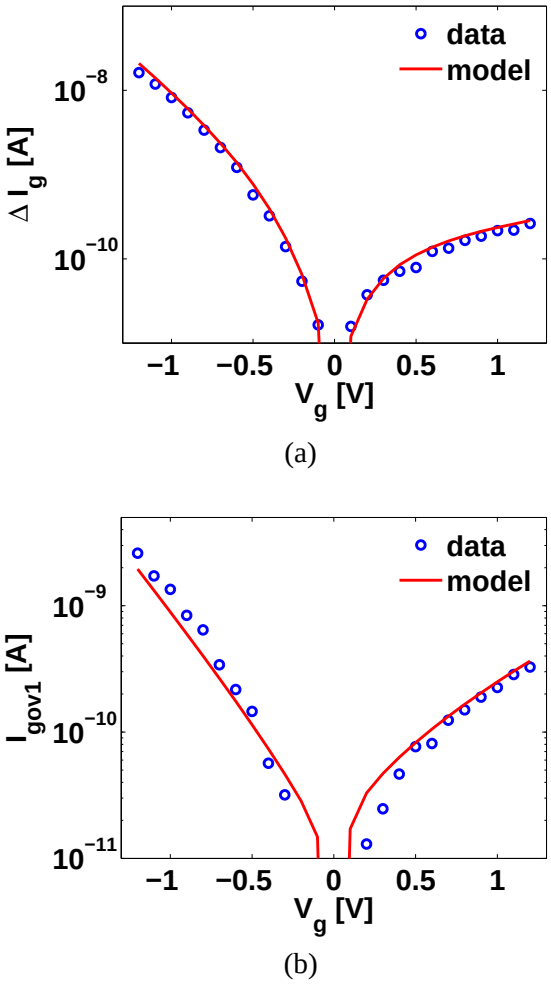



Fig. 8. (a) Measured and simulated $\Delta I_g(V)$ of (12) for $L_{gmax} = 0.5\ \mu m$, $L_{gmin} = 0.1\ \mu m$, and $W_g = 3\ \mu m$, and (b) Measured and simulated $I_{gov1}(V)$ of (13) for $L_{gmin} = 0.1\ \mu m$ and $W_g = 3\ \mu m$; devices used are in a 65nm RFCMOS technology.

to determine the parameters of the overlap region gate current model. At a given bias, the channel region gate current per unit length is given by $\Delta I_g/(L_{gmax} - L_{gmin})$, and this should be scaled by the effective length, see (2), to determine the intrinsic channel region current. This can then be subtracted from the total measured gate current of the minimum length device, $I_{g1}(V)$, to give the overlap region gate current

$$I_{gov1}(V) = I_{g1}(V) - \Delta I_g \cdot \frac{L_{gmin} + \text{DLQ}}{L_{gmax} - L_{gmin}} \ . \quad (13)$$

The parameters IGOVHVW and NOVO [1] can then be determined by fitting this current, see Fig. 8b.

The remaining parameters of the MOSVAR v1.0.0 model, including the parasitic resistances, are extracted as explained in [4].

Typical gate current fitting results are shown in Figs. 9 and 10 for a 130 nm CMOS process and in Figs. 11 and 12 for a 65 nm RFCMOS process. The good agreement between measurements and the model further validates our DLQ and DWQ extraction procedure presented in section II.

## IV. Conclusion

We have presented a new procedure to determine parameters of the MOSVAR varactor model, and have verified the technique by fitting $C(V)$ and tunneling current measurements over multiple geometries for several technology nodes.

## Acknowledgements

We would like to acknowledge I. Amano and R. Hiroyuki of Fujitsu Microelectronics Limited for providing the data for the devices in the 65nm technology node. This work was supported in part by the Compact Model Council.

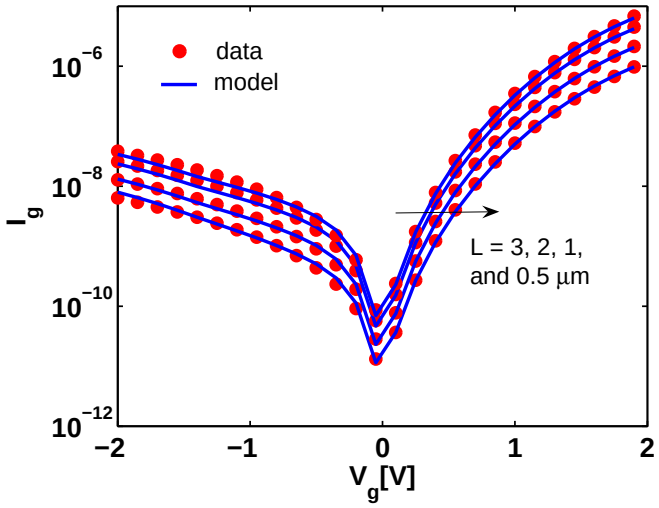



**Fig. 9.** Measured and simulated gate tunneling current for various $W_g$ devices in a 130 nm technology; $L_g = 0.5\ \mu m$ and $m = 20$.

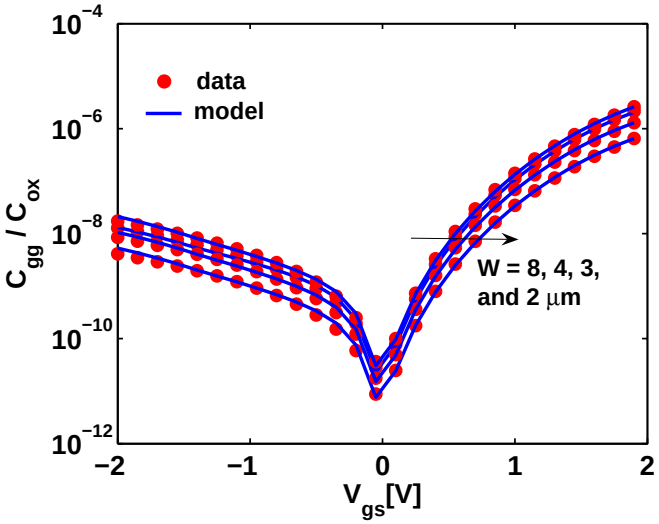



**Fig. 10.** Measured and simulated gate tunneling current for various $L_g$ devices in a 130 nm technology; $W_g = 3\ \mu m$ and $m = 20$.

## References

[1] J. Victory, Z. Zhu, Q. Zhou, W. Wu, G. Gildenblat, Z. Yan, J. Cordovez, C. McAndrew, F. Anderson, J.C.J. Paasschens, R. van Langevelde, P. Kolev, R. Cherne, and C. Yao, "PSP-based scalable MOS varactor model," *Proc. IEEE CICC*, pp. 495-502, Sep. 2007.

[2] 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 Trans. Electron Dev.*, vol. 53, no. 9, pp. 1979-1993, Sep. 2006.

[3] F. Li, L.F. Register, M.M. Hasan, and S.K. Banerjee, "A program for device model parameter extraction from gate capacitance and current of ultrathin SiO2 and high-K gate stacks," *IEEE Trans. Electron Dev.*, vol. 53, no. 9, pp. 2118-2127, Sep. 2006.

[4] J. Victory, Z. Yan, G. Gildenblat, C. McAndrew, and J. Zheng, "A physically based, scalable MOS varactor model and extraction methodology for RF applications," *IEEE Trans. Electron Dev.*, vol. 52, no. 7, pp. 1343-1353, Jul. 2005.

[5] J. Victory, C. McAndrew, and K. Gullapall, "A time-tependent, surface potential based compact model for MOS capacitors," *IEEE Electron Device Letters*, vol. 22, no. 5, pp. 245-247, May 2001.

[6] R. Tsu and L. Esaki, "Tunneling in a finite superlattice," *Applied Physics Letters*, vol. 22, pp. 562-564, 1973.

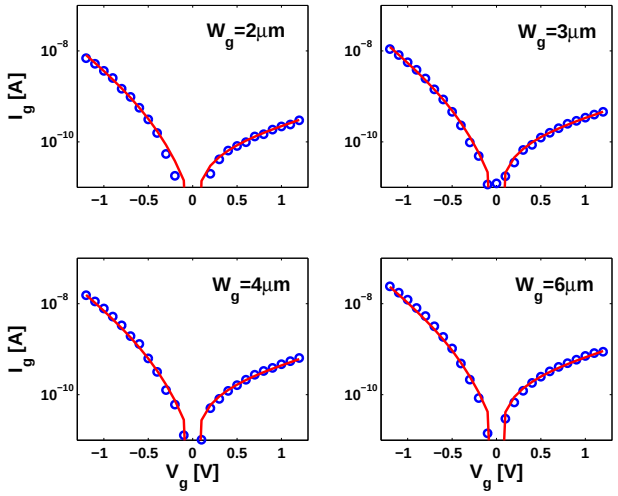



**Fig. 11.** Measured (oen circles) and simulated (solid line) gate tunneling current for various $W_g$ devices in a 65nm RFCMOS technology; $L_g = 0.2$ $\mu m$ and $m = 45$.

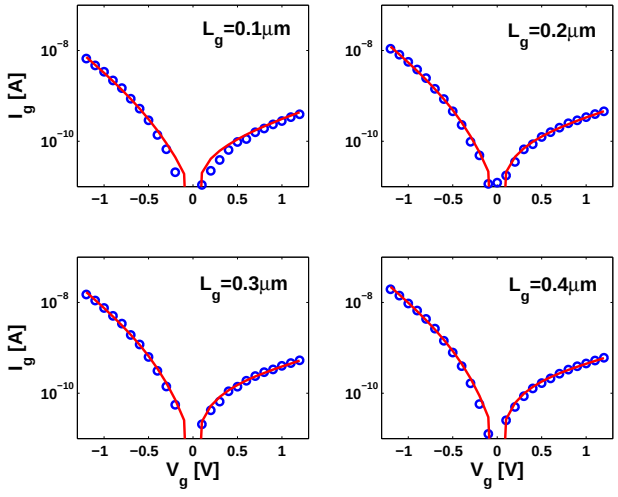



**Fig. 12.** Measured (open circles) and simulated (solid line) gate tunneling current for various $L_g$ devices in a 65nm RFCMOS technology; $W_g = 3$ $\mu m$ and $m = 45$.

# Nanomechanical test structure for optimal alignment in stencil-based lithography

Marc Sansa[1], Julien Arcamone[1], Jaume Verd[2], Arantxa Uranga[2], Gabriel Abadal[2], Emilio Lora-Tamayo[1], Núria Barniol[2], Marc A. F. van den Boogaart[3], Veronica Savu[3], Juergen Brugger[3], and Francesc Perez-Murano[1]

[1] Instituto de Microelectronica de Barcelona CNM-IMB (CSIC), Campus UAB, 08193 Bellaterra (Barcelona), Spain.
[2] Dept.Electronics Engineering, Universitat Autonoma de Barcelona, ETSE-UAB, 08193-Bellaterra (Barcelona), Spain
[3] Microsystems Laboratory (LMIS1), EPFL, EPFL, CH-1015 Lausanne (Switzerland).

## ABSTRACT

A nanoelectromechanical mass sensor based on a submicron size resonating metallic beam is used to characterize material deposition rates in stencil lithography. The material flux through micron size apertures is mapped with high spatial (below 1 μm) and deposition rate (below 10 pm/s) resolutions by displacing stencil apertures above the sensor. It is discussed how the sensor can be used as a test alignment for multi-level nanostencil lithography.

## INTRODUCTION

Shadow masking (also known as nanostencil lithography, SL) is a well known technique to fabricate patterns on a surface [1]. It is a versatile method that can be used in a variety of applications. There has been recently a strong interest regarding the use of shadow masks, mostly related to combinatorial materials science, organic based device fabrication, as well as rapid prototyping of nanoscale structures, using dynamic or quasi-dynamic stencil deposition. [2-9]

From the study of almost all reported variants of SL, a series of intrinsic generic advantages emerges. Its main features are its 'cleanliness', its flexibility, its parallelism and its high resolution. This exclusive characteristic makes that ultra-clean surfaces can be obtained with high purity deposits. From that, mechanically fragile and chemically functionalized surfaces can be structured, due to the absence of cyclic process steps seen in lithography and the absence of etching processes. Its parallelism makes it much faster than charged particle techniques (FIB or EBL). Recently it has been demonstrated its implementation at full-wafer scale, while providing 150 nm resolution. [10]

The issue of an alignment system for SL has not been addressed before. Having alignment capabilities enables the use of full wafer sized stencil and substrate, in combination with a large range and highly accurate XY nanopositioning stage, to perform stencil lithography in a quasi-dynamic approach [11]. The patterning procedure consists in aligning the stencil and the substrate within the vacuum chamber, putting them in contact, evaporating a material through the nanostencil apertures, stopping the deposition and separating them again, performing a controlled motion, aligning and evaporating again, eventually with another material. Repeating n-times this step-and-repeat approach, multiple nanometer scale patterns of multiple materials can potentially be obtained. Another decisive advantage of this technique is the opportunity to form high purity deposits, as the vacuum is not broken in between each nanopatterning step: very specific devices, like tunnel junctions, can be fabricated in this way.

We present the use of a dedicated sensor structure with spatial resolution, to test the alignment in stencil lithography systems. Specifically, we have employed a nanoelectromechanical mass sensor monolithically integrated in CMOS circuitry [12]. A nanoeletromechanical mass sensor is based on a resonating mechanical structure, which resonance frequency changes when a small quantity of material is deposited on it. Monitoring the change of resonance frequency, the deposition of mass can be monitored on line. The advantage of using this kind of test structure is that it allows performing the alignment detection using the flux of atoms through small apertures in the stencil membrane, i.e., sharing the same elements than in the normal stencil.

## EXPERIMENTAL SET-UP

The overall description of the system is shown in Fig. 1. Within the chamber of a metal evaporator (xv) a stencil membrane (vii) with the apertures defining the pattern to be transferred to the substrate (xiv) is fixed to a positioning X/Y/Z stage (v and vi) and to the stencil holders (xii and xiii). The stencil is located close to the substrate (or sample) (viii), where the pattern has to be realized by the material evaporated from the sources (ii and iii), through the patterning holes (xiv). The substrate is positioned on a substrate holder (ix). The alignment system is implemented by having additional apertures in the stencil (x) and the high sensitivity mass test structure co-planar with the substrate (xi). The structure detects when the alignment apertures in the stencil (x) are exactly on top of it, and so, a reference position can be set.

The most optimal situation is to have the sensor co-planar with the sample (viii), but, because of packaging limitations in the present configuration (which is intended as a proof of concept), it is located on the substrate holder. The reason is the bonding wires that connect the sensor to a printed circuit board. Advanced packaging methods are being developed to overcome this limitation.

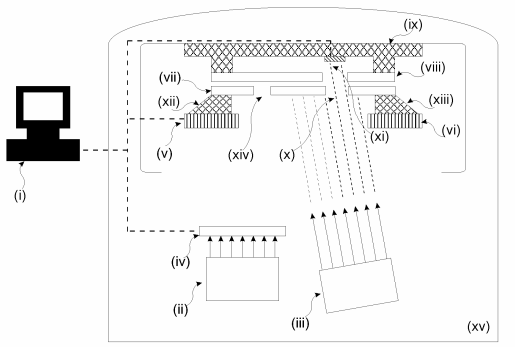



Figure 1. Diagram showing the overall stencil system with sensor alignment

## MODELING

The accuracy is related with the dimensions of the alignment apertures in the stencil and the dimensions of the sensor: the smaller they are, the most accurate will the position sensing be. When decreasing those dimensions, less signal (material) is arriving to the sensor, so that high sensitivity (mass) sensors are required. In the following, we model the spatial distribution of the deposition rate below the stencil as a function of the overall system dimensions.

In shadow masking, a purely geometrical dispersive effect leads to pattern widening when the vertical distance between stencil apertures and substrate, defined as the gap $G$, is not null. The resulting pattern is inherently wider than the nominal one which is defined by the aperture dimensions. The pattern widening $\Delta W$ on each side is given by:

$$\Delta W = \frac{1}{2}\frac{SG}{D} \approx 15\,\mu\mathrm{m} \qquad (1)$$

where $S$ is the source diameter (~ 1 cm) and $D$ is the source-sensor vertical distance (~ 50 cm). In our system, the gap $G$ is 1.5 mm because of the sensor packaging. This has two consequences: the positioning accuracy is potentially decreased, and the specifications on the sensor performance are more restricting due to the fact that a smaller material flux is received per unit area on the sensor plane, i.e. the sensor needs to be more sensitive. Indeed, if considering the stencil-substrate gap negligible respect to the substrate-sensor gap, the area covered by the material flux passing through a rectangular aperture and impacting on the sensor plane can be approximated as:

$$\begin{aligned} A_{SP} &= (W_{ST\_X} + 2\Delta W)(W_{ST\_Y} + 2\Delta W) \approx \\ &\approx (W_{ST\_X} + 30)(W_{ST\_Y} + 30)\mu m^2 \end{aligned} \qquad (2)$$

where $W_{ST\_X}$ and $W_{ST\_Y}$ are the effective aperture dimensions (in µm) in $X$ and $Y$ axes respectively. The term 'effective' refers to the fact that an aperture 'seen' by the sensor is actually the real-time intersection between the stencil aperture and the substrate one which is constantly located above the sensor. For small apertures, $A_{SP}$ is practically independent of $W_{ST}$.

In 3D, the corresponding deposition profile on the sensor plane is a truncated pyramid (i) whose inferior basis has an area given by (2), and (ii) whose superior basis represents the area $A_{MAX}$ of maximum flux $F_{SM}$. Our geometrical model predicts that when $W_{ST} > 2\,\Delta W$ (see Fig.2), $F_{SM}$ is equal to the nominal source flux $F_{NS}$. (i.e. the maximum relative flux, $F_{RM}$= 1), $A_{MAX}$ being given in this case by:

$$(W_{ST\_X} - 2\Delta W)(W_{ST\_Y} - 2\Delta W) \qquad (3)$$

However, when $W_{ST} < 2\,\Delta W$ (see Fig. 2), which is the case in our system, and more generally when small apertures are used with large gaps, an important issue is that the deposition rate on the sensor plane becomes smaller than the nominal source flux according to **Error! No s'ha trobat l'origen de la referència.** because of the previously mentioned "penumbra" effect. $A_{MAX}$ is then given by :

$$(2\Delta W - W_{ST\_X})(2\Delta W - W_{ST\_Y}) \qquad (4)$$

The quantities of mass related to these material fluxes impacting on the sensor are extremely small and require high-performance mass sensors. Let us roughly estimate the quantity of mass $\Delta m$ deposited during the motion of a stencil square aperture (of width $W_{ST}$) above a substrate square aperture (of width $W_{SUB}$)

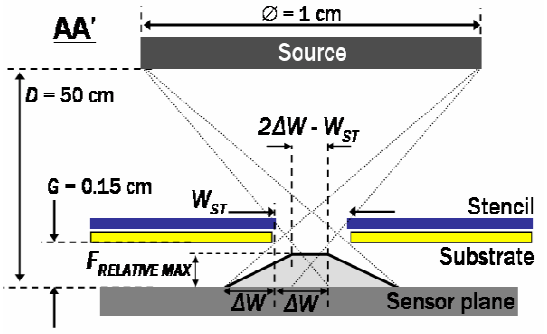



Figure 2. Pattern widening effect as a function of the geometrical dimensions of the system, for $W_{ST} < 2\,\Delta W$

below which a QCM is located (whose size largely exceeds these apertures):

$$\Delta m \approx \rho_D W_{ST}^2 W_{SUB} \frac{F_S}{v_{scan}} \qquad (5)$$

where $F_S$ is the material flux impacting on the sensor plane, $v_{SCAN}$ is the scanning speed and $\rho_D$ is the density of the deposition material. Then, the resulting frequency shift $\Delta f$ can be expressed as a function of the punctual mass sensitivity of the QCM $S_{PUNCTUAL\ QCM}$ (the area covered by the flux on the QCM plane is so small compared to its own area that one can consider the QCM operated in punctual mode):

$$\Delta f \approx \frac{\Delta m}{S_{PUNCTUAL\_QCM}} = \frac{\rho_D W_{ST}^2 W_{SUB}}{S_{PUNCTUAL\_QCM}} \frac{F_S}{v_{scan}} \qquad (6)$$

The numerical calculation with realistic values of this system ($W_{ST}$ = 20 µm; $W_{SB}$ = 40 µm; $\rho_D$ = 10.5 g.cm$^{-3}$; $F_S$ = 0.2 nm.s$^{-1}$; $v_{SCAN}$ = 3 µm.s$^{-1}$ and $S_{PUNCTUAL\ QCM} \approx$ 1.8 10$^{-8}$ g.Hz$^{-1}$) leads to $\Delta f \approx$ 0.6 mHz. This value is approximately one order of magnitude smaller than the best possible frequency resolution of a QCM, in other words the performance of QCM is too poor to make this type of mass measurements.

## NANOELECTROMECHANICAL MASS SENSOR

The electromechanical mass sensor structure is a 18 µm long, 600 nm wide, 850 nm thick (nominal dimensions) resonant doubly clamped beam, whose metallic structure is fabricated with the top metal layer of a commercial 0.35 µm CMOS technology (see Fig. 3).

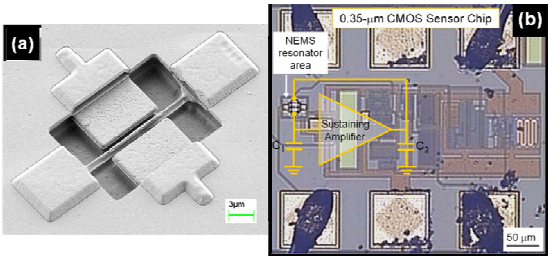



Figure 3. SEM image of the Nanomechanical resonator (left) and optical image showing the sensor + CMOS circuit

Two lateral electrodes, placed at 600 nm from the resonator on both sides, are used for electrostatic actuation and capacitive readout. The electromechanical resonator is monolithically integrated with a CMOS oscillator circuit. The resonance frequency of the CMOS-NEMS is around 14 MHz, with a quality factor around 1000 in vacuum.

The mass sensitivity of this device has been calibrated by depositing a 2 nm thick silver layer, indicating an areal mass sensitivity of 3.4×10$^{-11}$ g.cm$^{-2}$.Hz$^{-1}$. The resonator is electrostatically self-excited so that it sets the oscillation frequency of the entire oscillator circuit [13]. The frequency stability level (measured in terms of Allan deviation), combined with the device mass sensitivity, leads to a mass resolution of $\delta M$ = 450 pg.cm$^{-2}$, corresponding to a thickness resolution of 0.43 pm for Ag deposition. As a result, the NEMS sensor structure used here presents the following specific features: (i) a high mass resolution which allows detecting deposition rates below 10 pm.s$^{-1}$ for Ag deposition, better than state-of-the-art quartz crystal microbalances (QCM); (ii) its small size intrinsically provides a spatial resolution in the range of hundreds of nm which allows a position-dependent detection; (iii) device portability, because the sensor actuation and readout are completely electrical (CMOS integration).

## EXPERIMENTAL RESULTS

Fig. 4 shows the change of frequency when the stencil is scanning and a hole crosses over the sensor. Its derivative directly corresponds to the instantaneous and spatially resolved deposition rate.

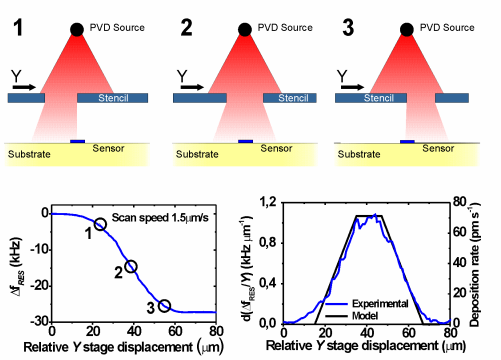



Figure 4. Experimental determination of the evaporation rate of an atom beam after passing through an stencil hole

In the present implementation, a dense array of micromachined square apertures (40 x 40 µm$^2$) is defined into the substrate to make possible the manual pre-positioning of one of these apertures above the mass sensor without any extra-assistance. Fig. 5a shows an optical image of the sensor after more than thirty depositions: the result of successive depositions through all the surrounding substrate alignment-dedicated holes can be observed. The sensor is remarkably located inside the planar projection of a substrate aperture onto the sensor plane, thus confirming a correct initial positioning between both tests.

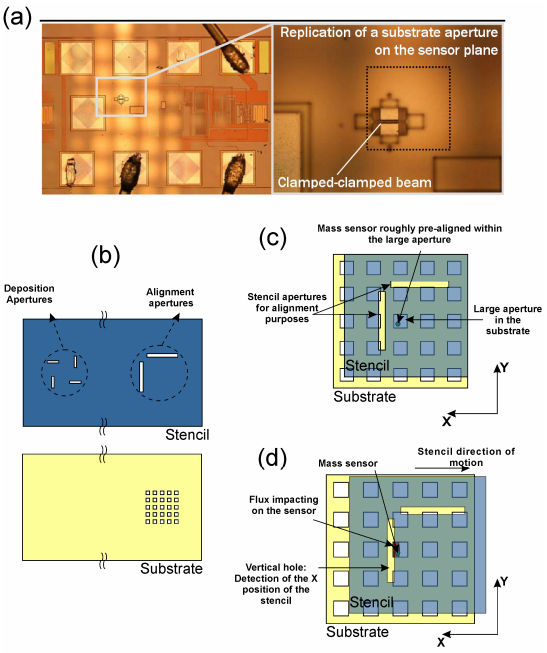



Figure 5. Alignment apertures in the substrate and stencil.

The apertures on the stencil are designed as two orthogonal rectangles (slits) (Fig. 5b), one each for aligning in the *X* and *Y* axes respectively. The stencil also contains another set of apertures which are specifically dedicated to the definition of patterns on the substrate.

To estimate the angular error, we propose the following alignment procedure: it consists in successively displacing the stencil along the *X*-axis but at two different *Y* positions. Fig. 6 shows schematizes the change of resonance frequency of the NEMS mass sensor when the sensor is displaced below the area of an alignment-dedicated stencil mark. The variation of the *X* position, $\Delta X$, where the change of resonance frequency occurs is related to the relative angle between the stencil pattern and the *XY* axis of the stage. Knowing $\Delta X$ and $\Delta Y$ (*Y* step between both *X* scans), the angular error can be determined and further compensated by software.

$$\theta = \arctan\frac{\Delta X}{\Delta Y} \qquad (7)$$

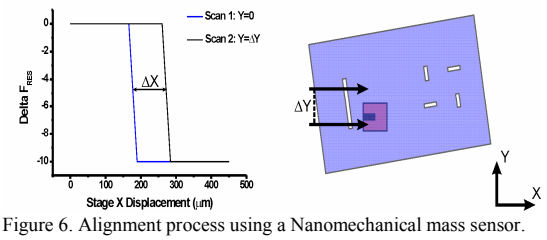



Figure 6. Alignment process using a Nanomechanical mass sensor.

The alignment resolution of this system is estimated from the configuration shown of Fig. 5.b to d. We consider the resolution to be the uncertainty in the determination of the mid-point of the transition zone of frequency shift (i.e. wherever $d\Delta f_{RES}/dX \neq 0$). It approximately stands in the range of 1 µm. We can define the angular misalignment resolution as the minimum angle the system can correct. This parameter is limited by the resolution of the system, that is, the minimum $\Delta X$ we can detect. Therefore we can estimate the angular misalignment resolution from (7). Using our current system resolution (1µm) and considering an alignment aperture ($\Delta Y$) of 400µm, the angular misalignment resolution is 0.143 degrees.

Alignment between stencil and XY movement axes can currently be achieved using SEM [11]. Fig. 7.a shows the result of three successive shadow-masked depositions (marked as red, blue and black in the optical image) without pre-alignment of the stencil with respect to the *XY* axes of the stage. The stencil pattern consists of four rectangular slits 100 x 4 $\mu m^2$ with a pitch of 800 µm. The second and third depositions are performed after translating the stencil respectively 100 µm and 800 µm in the *X* direction. The result (see the zoom in Fig. 7a) is a disconnected pattern due to an angular misalignment.

After de first deposition sequence, the substrate is inspected using SEM in order to determine the angular misalignment. This allows correcting the XY movement, therefore compensating the misalignment. Fig. 7.b shows the second deposition sequence after applying the angular correction: the resulting patterns are correctly positioned and connected.

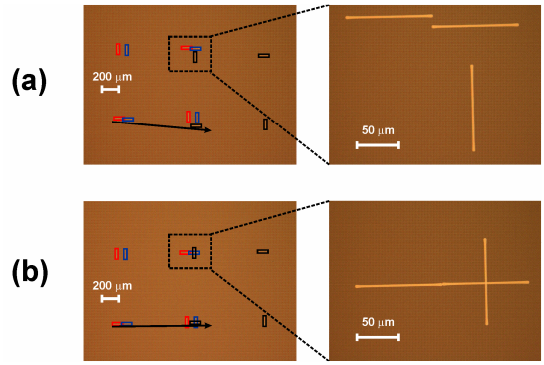



Figure 7. Example of alignment performed using SEM Microscope to correct the stencil misorientation and alignment process using a Nanomechanical mass sensor. (a) Patterns misaligned due to stencil movement misorientation. (b) Second deposition sequence after measuring and correcting misalignment using SEM. After the correction, the measured misalignment angle is 0.15 degrees.

After the second deposition sequence, the misalignment angle was still found to be 0.15 degrees. This value is similar to the angular misalignment error that can be provided by the mass sensor method, even though under a conditions where the experimental set-up is not optimum.

Regarding future perspectives, we have theoretically studied the ultimate alignment resolution limits using the geometrical model. For this purpose, the minimum aperture size $W_{ST\ MIN,}$ that still would be sensed is evaluated for different sensor lengths $L_S$. In the submicron range or below, we actually consider $W_{ST\ MIN}$ as the alignment resolution itself. $W_{ST\ MIN}$ has been calculated as a function of the stencil-sensor gap $G$ for a minimum detectable deposition rate $F_{S\_MIN}$ of 0.01 nm/s (see additional information part II), which has been experimentally observed for $L_S$ = 18 µm. In future works, special efforts will be dedicated to the sensor packaging in order to reduce $G$. In the ultimate case, when $G$=0, it is found that:

$$W_{STMIN}(G=0)=\frac{F_{S\_MIN}}{F_{NS}}L_S \qquad (8)$$

assuming that $L_S$ is equal or superior to the sensor width, and that the approximation of a constant areal mass sensitivity is still valid even for small areas. Considering a constant scanning speed of 3 µm/s, we extrapolate $F_{MIN}$ for more sensitive resonators of smaller length that are still feasible to be both fabricated and sensed (see for example [9]). Given that the sensor sensitivity $S_A$ scales as the square of $L_S$ $^{6}$, $F_{MIN}$ also scales with the square of $L_S$. Therefore, $W_{ST\ MIN}$ for $G$ = 0 (see Eq. (8)) scales as $L_S$ $^{3}$. As shows Fig. 8, 100 nm apertures (i.e. alignment resolution) are likely to be detected just by reducing the sensor length by a factor of 2. Furthermore, the angular misalignment resolution improves linearly with the alignment resolution if we consider it to be small, as seen in (7).

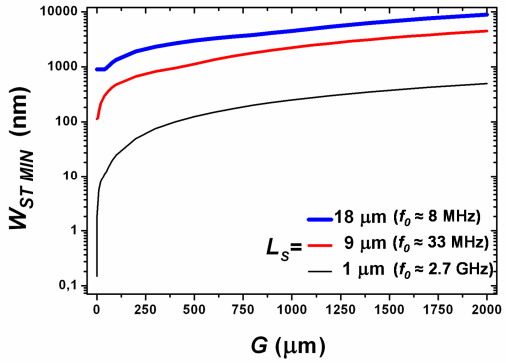



Figure 8. Evaluation of the minimum detectable stencil aperture $W_{ST\ MIN}$ as a function of the stencil-sensor gap G for a relative minimum detectable flux of 0.05 and a scan speed of 3 µm/s

## CONCLUSIONS

We have presented a concept to align a mask and a substrate in order to perform sequential 'in situ' patterning by shadow masking. The alignment system is based on the spatially-resolved detection of dedicated apertures of the stencil membrane by using a nanomechanical mass sensor. The concept has been demonstrated by correcting the misorientation of a stencil pattern with respect to the displacement axes of the stage. We experimentally show that the nanomechanical mass sensor allows mapping the spatial distribution of the deposition rate below the sensor, and the results are found to be in agreement with a purely geometrical modeling of the system. This model allows us to infer the minimum aperture that can be detected, and in consequence, the alignment resolution. Our analysis tends to conclude on the capability of sub-100 nm alignment in future developments only requiring light modifications. More generally, this work opens new perspectives for fabricating complex nanoelectronic devices with clean interfaces.

## ACKNOLWEDGMENTS

The work is partially funded by projects TEC2006-03698/MIC, TEC2006-13910-C03, FP6 EU-project NaPa

## REFERENCES

[1] J. Brugger, V. Savu, K. Sidler, M.A.F. van den Boogaart, O. Vazquez Mena, G. Villanueva, *E-Nano Newsletter*, vol 8, pp 22-28, June 2007

[2] A. Ludwig, J. Cao, J. Brugger et al., Journal of Combinatorial Chemistry 6 (1), 50-53 (2004).

[3] A. Tixier, Y. Mita, J.P. Gouy, H. Fujita; "A Journal of Micromechanics Microengineering; Vol.10; pp. 157-162; (2000)

[4] P. F. Tian, P. E. Burrows, and S. R. Forrest Applied Physics Letters 71 (22), 3197-3199 (1997);

[5] Y. X. Zhou and A. T. Johnson, Nano Letters 3 (10), 1371-1374 (2003).

[6] S. Egger, A. Ilie, Y. T. Fu et al., Nano Letters 5 (1), 15-20 (2005);

[7] L. Gross, R. R. Schlittler, G. Meyer et al., Applied Physics Letters 90 (9) (2007);

[8] H. M. Guo, D. Martrou, T. Zambelli et al., Applied Physics Letters 90 (9) (2007).

[9] Z. Racz, J. L. He, S. Srinivasan et al. Journal of Vacuum Science & Technology B 22 (1), 74-76 (2004).

[10] J. Arcamone, M.A.F. van den Boogaart, F. Serra-Graells,; J. Fraxedas, J. Brugger, F. Pérez-Murano; Nanotechnology, 19, 305302 (2008)

[11] V.Savu, M.A. F. van den Boogaart, J.Brugger, J. Arcamone, M. Sansa, and F. Perez-Murano. J.Vac.Sci.Technol.B, 26 2054 (2008)

[12] J.Verd, A.Uranga, J.Teva, J.L.López, F.Torres, J.Esteve, G.Abadal, F.Pérez-Murano, N.Barniol IEEE Electron Device Letters ,27, 495-497 (2006)

[13] J. Verd, J. Verd, A. Uranga, G. Abadal, J. Teva, F. Torres, F. Perez-Murano, J. Fraxedas, J. Esteve, N. Barniol Applied Physics Letters, 91, 013501 (2007)

# Electrical Test Structures for Investigating the Effects of Optical Proximity Correction

A. Tsiamis*, S. Smith*, M. McCallum†, A.C.Hourd‡, J.T.M. Stevenson* and A.J. Walton*

* Institute for Integrated Micro and Nano Systems, School of Engineering,
Scottish Microelectronics Centre, The University of Edinburgh, Kings Buildings, Edinburgh, EH9 3JF, U.K.
Email: A.Tsiamis@ed.ac.uk

†Nikon Precision Europe GmbH, Appleton Place, Appleton Parkway, Livingston, West Lothian, EH54 7EZ, U.K.

‡Compugraphics International Ltd., Eastfield Industrial Estate, Glenrothes, Fife, KY7 4NT, U.K.

## ABSTRACT

Electrical test structures have been designed to enable the characterisation of corner serif forms of optical proximity correction. These structures measure the resistance of a conducting track with a right angled corner. Varying amounts of OPC have been applied to the outer and inner corners of the feature and the effect on the resistance of the track investigated. A prototype test mask has been fabricated which contains test structures suitable for on-mask electrical measurement. The same mask was used to print the structures using an i-line lithography tool for on-wafer characterisation. Results from the structures at wafer level have shown that OPC has an impact on the final printed features. In particular the level of corner rounding is dependent upon the dimensions of the OPC features employed and the measured resistance can be used to help quantify the level of aggressiveness of the inner corner serifs.

## INTRODUCTION AND BACKGROUND

The work presented in this paper is part of an ongoing project which uses electrical test structures for the characterisation of advanced photomasks. Previous publications have demonstrated that on-mask test structures are capable of fast and repeatable sheet resistance and critical dimension (CD) measurements on binary and alternating aperture phase shifting masks (Alt-PSMs) [1], [2], [3], [4]. More recently, industry standard optical metrology patterns have been adapted into electrical equivalents on binary masks. These have been used to investigate iso-dense proximity effects [5], to characterise the GHOST proximity correction technique [6] and to compare different metrology methods [7]. The fabricated masks also contained test structures to investigate the dimensional mismatch between closely spaced chrome features [8]. A plate which contained a large number of these structures was subsequently fabricated and this provided an unprecedented volume of data on dimensional mismatch [9] in a photomask manufacturing process.

This paper attempts to extend the characterisation of proximity correction techniques, by investigating the effect of optical proximity correction (OPC) in the form of two-dimensional corner serif features. Corner serifs are frequently used in advanced photomask designs and present a challenge to conventional metrology techniques [10]. Electrical test structures have been designed for this work and initial results from measurements made on a prototype binary mask (MSN6754) have demonstrated that it is feasible to electrically characterise this method of OPC [11] on mask. Although practically restricted to inner corner serifs, this is not considered a problem, as it is this circuit feature that is most relevant to achieving device performance. Furthermore, inner corner serif structures are the most difficult to manufacture and quantify using conventional optical metrology techniques. For the first time, the photomask, which was also measured electrically has been used to print test structures on wafer. Initial, on-wafer, measurements of these structures have been made and an analysis of the results is reported here.

## TEST STRUCTURES AND FABRICATION

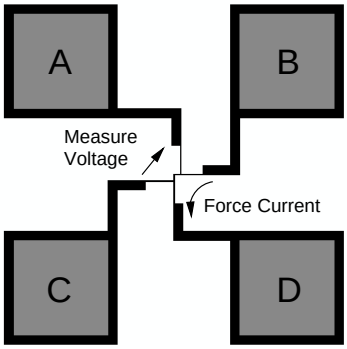
The designed test structure is a Kelvin connected resistor consisting of a short section of metal track turning through a right angled corner, as shown in Fig. 1(a). To measure the resistance of the track a current is forced between pads B and D and the resulting potential difference is measured at pads A and C. The resistance of the section of track between the voltage taps is then calculated with

$$R = \frac{V_{AC}}{I_{BD}} \tag{1}$$

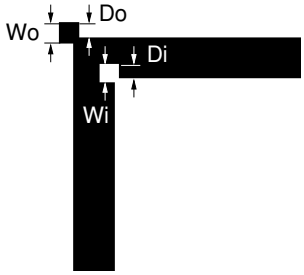
The effect of OPC is investigated by altering the layout of the right angled corner of the structure. Specifically, a square of metal is added to the outside of the corner while a square is removed from the inside. This is illustrated in Fig. 1(b). The OPC aggressiveness is altered by changing the size of the square ($W_i$ or $W_o$) and the amount by which it overlaps with, or protrudes from, the original layout ($D_i$ or $D_o$).

The test mask design has test structures with 3 different base values of CD: 1.6$\mu$m, 2.0$\mu$m and 2.4$\mu$m. These correspond to printed dimensions of 320nm, 400nm and 480nm when imaged with the 5$\times$ projection lithography tool used

978-1-4244-4259-1/09/$25.00 ©2009 IEEE




(a) Kelvin connected resistor structure




(b) Expanded view of the structure

Fig. 1. Layout of Kelvin test structure and closeup showing parameters of OPC features.

for this work. The dimensions of the OPC elements are defined as fractions of the base CD: 0.25, 0.3, 0.35, 0.4, 0.45 and 0.5. Subsequently, the value of $D_i$ or $D_o$ is then defined as some fraction of $W_i$ or $W_o$: 0.25, 0.5 or 0.75. The test mask has the full range of OPC dimensions for the 1.6$\mu$m (320nm on-wafer) structures and a reduced set for the other CDs with only the $D_i$ and $D_o$ values set to half of $W_i$ or $W_o$.

On-wafer test structures were fabricated in a 300nm thick layer of doped polysilicon deposited on a 0.5$\mu$m thick layer of thermal silicon dioxide on 200mm silicon substrates. The wafers were spin coated with Ultra-i 123 i-line photoresist and printed using a 5$\times$ Nikon i-line step and repeat lithography tool. After development the polysilicon was etched in a reactive ion etch tool before passivation with a 0.5$\mu$m thick layer of PECVD silicon oxide. Holes in the oxide were then etched over the probe pads before deposition of a 0.5$\mu$m thick layer of sputtered aluminium. The final step was to pattern the aluminium to create contacts suitable for probing.

## MEASUREMENTS AND RESULTS

### A. Electrical Measurements

The results from the measurements made on the on-mask test structures have shown that the electrical technique is sensitive enough to measure the effects of inner corner structures reliably and in good agreement with the simulated predictions [11]. The agreement with the simulations, presented in Fig. 2, is a good indicator that it is the chrome material present on the mask which is being characterised, removing some of the ambiguities inherent when interpreting indirectly acquired images of the mask pattern. Any departures from the simulations have been found to be attributable to defects in the serif structure. This can be seen in Fig. 3 as a measurement point which deviates from the anticipated trend of the curves. It is now hoped that the printed structures and the electrical measurement methods applied so far, will form a valuable tool for investigating the quality of lithographic transfer and for optimising the corner serif structures.

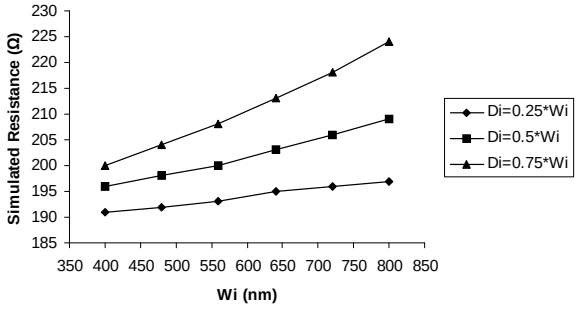



Fig. 2. On-mask resistance simulations from structures with inner corner OPC features.

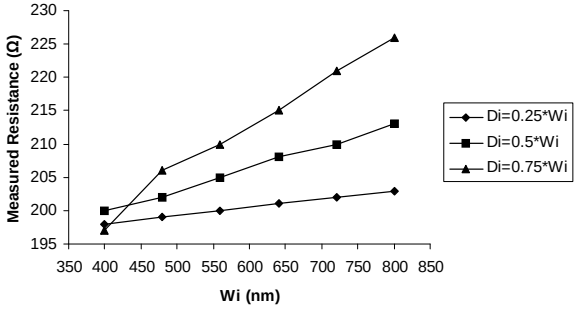



Fig. 3. On-mask electrical measurement results from structures with inner corner OPC features.

Three sets of on-wafer structures with varying nominal OPC feature dimensions were measured electrically and the results are presented in Fig. 5. One set consists of structures with inner corner OPC serifs only, one with outer corner serifs only and one with both. The resistance results of Fig. 5(a) suggest an upward trend in resistance as $W_i$ increases, as would be expected. However, the nature of the results suggest that other factors also affect the resistance

of the structure. The reason behind this could be related to fabrication artifacts such as non-uniform lithography or etching, which conceal the data trends because the resistance changes are not only due to the OPC features. Another factor that could strongly affect the results is any variability in the localised sheet resistance of the polysilicon layer. This could introduce some of the structure-to-structure variations shown in the resistance results. Reference [12] examines the effect that the geometry and grain structure has upon the value of resistivity extracted from a Greek cross structure. The sheet resistance ($R_S$) of a polysilicon film is a function of the size and distribution of the grains and [12] identifies how the grain structure increases the variability of $R_S$ measured using Greek crosses with narrow arm widths. Unfortunately there are no structures such as these on the wafer to confirm whether this is the cause of the variability observed in the measured results. A wafer map of the sheet resistance of the polysilicon layer, measured using a four point probe system on a wafer from the same batch as the test structures, is presented in Fig. 4. Unfortunately, this measurement does not have sufficient resolution to identify local sheet resistance variations that might cause the observed variation in the resistor test structures.

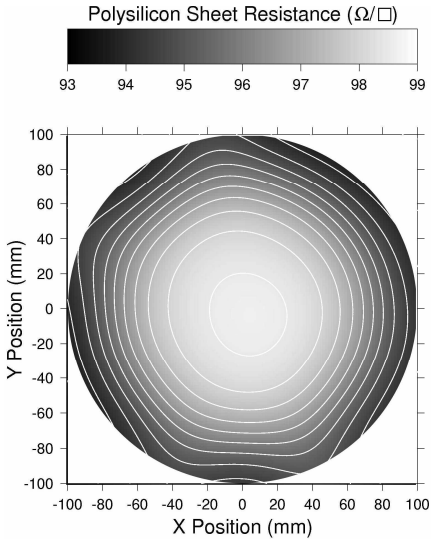
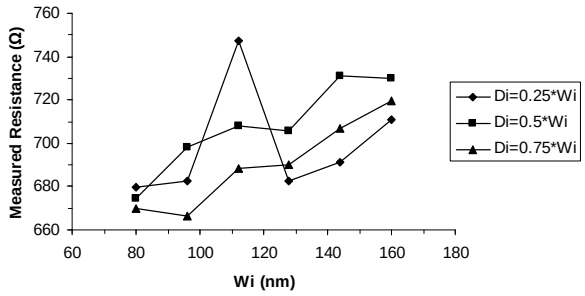
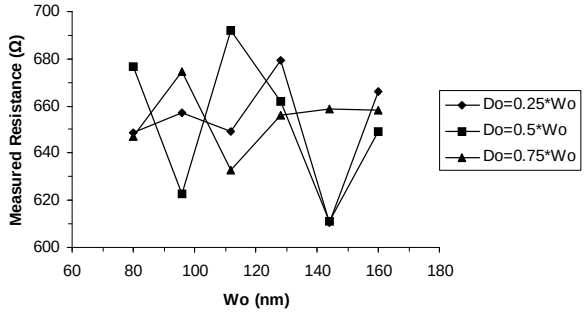



Fig. 4. Sheet resistance variation across one wafer with corner OPC electrical test structures.

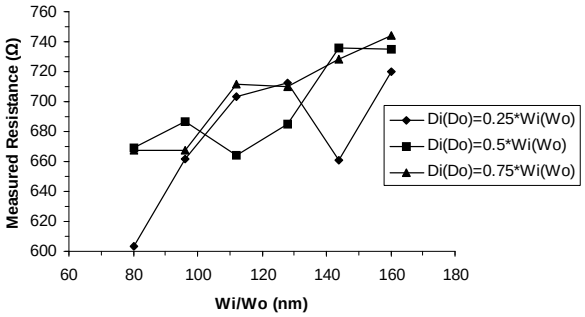
Fig. 5(b) indicates that the presence of OPC has little or no effect on the resistance of the conductive track when applied to the outside of the right angled corner. This behaviour is to be expected as most of the current flow in the structure is concentrated around the region of the inner corner. This is illustrated in Fig. 5(c) which shows results from structures with both inner and outer corner serifs. It is clear that the dominant effect in the measured resistance is from the inner corner section, as a similar trend is observed for structures with only inner corner serifs. These results are obviously also affected by significant resistance variation. Unfortunately only small numbers of the test structures have been tested so far and the inter-die variation has yet to be determined.




(a) Measured resistance against inner corner OPC dimensions




(b) Measured resistance against outer corner OPC dimensions




(c) Measured resistance against inner and outer corner OPC dimensions

Fig. 5. Resistance measurements for structures with inner and outer corner OPC arrangements.

### *B. Scanning Electron Microscope (SEM) Inspection*

In order to visually investigate the effects, a Philips XL40 SEM was used to capture images of the structures which were measured electrically. Imaging software was then used to invert the SEM images and filter out their background. Fig. 6(a) shows a structure with no OPC, while Fig. 6(b) and 6(c) show structures with the most aggressive inner

and outer corner serif dimensions respectively. The images suggest that when OPC is applied to the inner or the outer corner of a structure the shape of the corner changes and in particular the level of corner rounding.

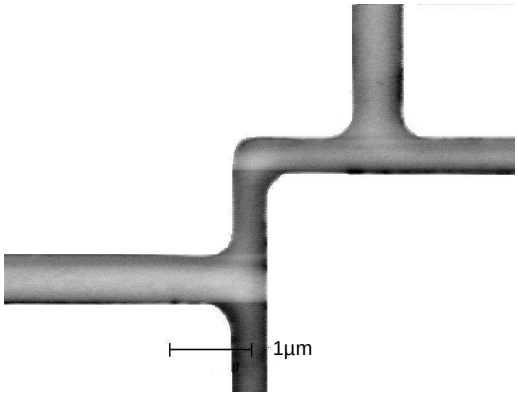
To help compare the structures an edge detection filter in the GNU Image Manipulation Program (GIMP) was applied to the images which are subsequently overlaid on one another. Comparisons of the structure with no OPC with the structures with the most aggressive OPC on the inner and outer corners are presented in Fig. 7. These confirm that when OPC is applied the shape of the inner and outer corners changes with respect to a uncorrected corner. Therefore, for inner corners OPC does have an effect and is likely to be the cause of the observed trends in electrical measurement results. On the other hand while OPC does affect the rounding of the outer corner it has no effect on the measured resistance and could be omitted from designs in many cases.

## SIMULATIONS

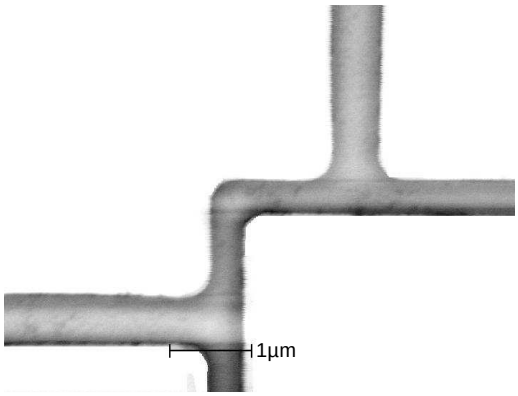
The two-dimensional (2D) solver of the interconnect analysis software *Raphael* was used to model the resistance of test structures with different levels of corner rounding applied to the inner or outer corner. A sheet resistance of $97\Omega/\square$ was chosen for the material of the simulated structures, as this represents a typical value for the polysilicon taken from the four point probe results shown in figure 4.

The results of the simulations for test structures with inner corner rounding are presented in Fig. 8(a), while those for structures with outer corner rounding can be seen in Fig. 8(b). The results of Fig. 8(a) show that when the there is a significant resistance change with respect to the inner corner rounding. In particular, it appears that the resistance strongly depends on the radius of the corner rounding and thus the area of the material added to the structure. This behaviour is to be expected as most of the current flow in the structure is concentrated around the region of the inner corner. This confirms that for the inner corner measurement results of Fig. 5(a), OPC has an effect on the fabricated corner of the printed structure, which is dependent to the dimensions of the OPC square.

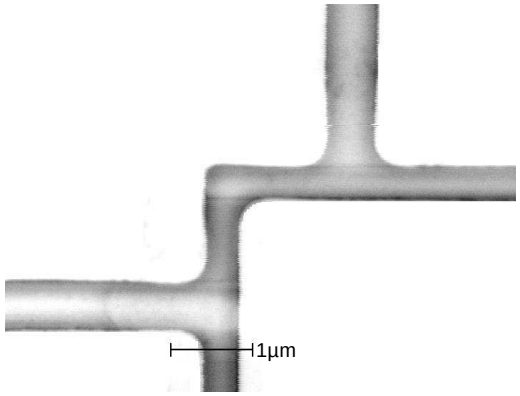
The simulations for outer corner rounding on Fig. 8(b) confirm that there is little variation of resistance when the area of the outer corner changes and this is to be expected as there is little or no current flow in this region of the structure. Any resistance variation caused by altering the dimensions of the radius of the outer corner rounding will be minimal compared to other fabrication effects on wafer. This can be seen on the results of Fig. 5(b) which appear noisy and show no trend with respect to dimensions.




(a) Image for a structure with no OPC




(b) Image for a structure with inner corner OPC

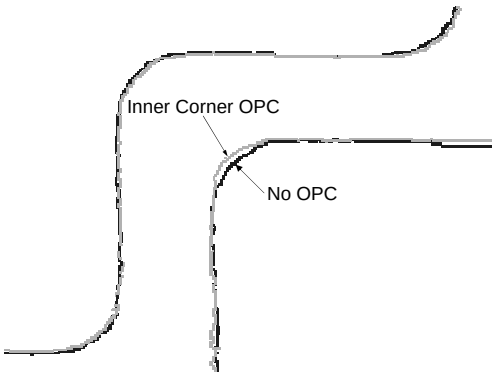



(c) Image for a structure with outer corner OPC

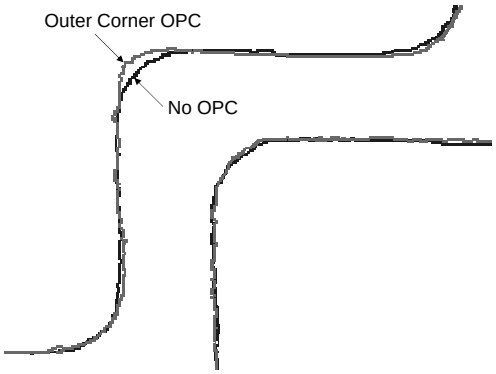
Fig. 6. SEM images of structures with different OPC arrangements.

## CONCLUSIONS

Electrical test structures have been designed to characterise OPC in the form of corner serifs and to investigate the impact they have on printed features. Results from on-mask structures have shown that the electrical measurement technique is sensitive enough to detect the effects of OPC




(a) Shapes of structures with inner corner OPC or no OPC

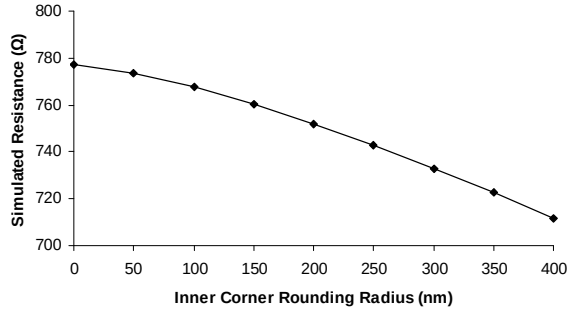



(b) Shapes of structures with outer corner OPC or no OPC

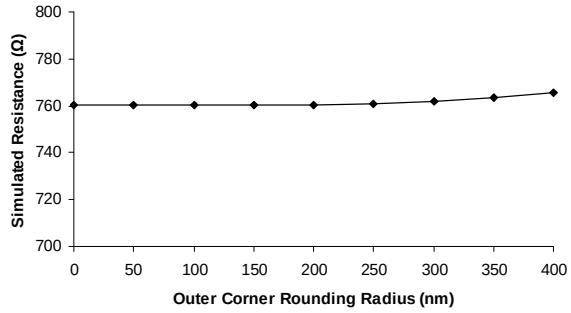
Fig. 7. Overlaid shapes of structures with different OPC arrangements.

on inner corners and any abnormalities on the fabricated features. In addition the presence of OPC has little effect on the measured resistance when applied to the outside of the corner structure.

Results of electrical measurements of polysilicon test structures printed using this photomask suggest that OPC applied to the inner corner has a significantly greater effect on the resistance than outer corner serifs. SEM imaging of the test structure geometries confirms that OPC does alter the shape of corner rounding. Therefore, it is primarily the inner corner OPC that affects the resistance of a measured structure, while outer corner OPC has little or no effect on the measurement. However, it appears that, unlike measurements on the mask, the effects of OPC are confounded by other fabrication artifacts when printed with i-line technology. Local variations cannot be explained by dimensional effects and the most likely explanation is the grain structure of the polysilicon, which is a well known cause of sheet resistance variation.




(a) Simulated resistance against inner corner rounding dimensions




(b) Simulated resistance against outer corner rounding dimensions

Fig. 8. Simulated resistance for structures with inner and outer corner OPC arrangements.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Edinburgh Research Partnership in Engineering and Mathematics, Nikon Precision Europe GmbH and Compugraphics International Ltd.

## REFERENCES

[1] S. Smith, M. McCallum, A. Walton, J. Stevenson, and A. Lissimore, "Comparison of Electrical and SEM CD Measurements on Binary and Alternating Aperture Phase-Shifting Masks," *IEEE Transactions on Semiconductor Manufacturing*, vol. 16, no. 2, pp. 266–272, May 2003.

[2] M. McCallum, S. Smith, A. Hourd, A. Walton, and J. Stevenson, "Cost Effective Overlay and CD Metrology of Phase-Shifting Masks," in *Proceedings of the SPIE, Vol. 5567: 24th Annual BACUS Symposium on Photomask Technology*, 2004, pp. 596–603.

[3] S. Smith, M. McCallum, A. Walton, J. Stevenson, P. Harris, A. Ross, and L. Jiang, "On-Mask CD and Overlay Test Structures for Alternating Aperture Phase Shift Lithography," *IEEE Transactions on Semiconductor Manufacturing*, vol. 18, no. 2, pp. 238–245, May 2005.

[4] S. Smith, A. Tsiamis, M. McCallum, A. Hourd, J. Stevenson, and A. Walton, "Comparison of Optical and Electrical Techniques for Dimensional Metrology on Alternating Aperture Phase-Shifting Masks," *IEEE Transactions of Semiconductor Manufacturing*, vol. 21, no. 2, pp. 154–160, May 2008.

[5] A. Tsiamis, S. Smith, M. McCallum, A. Hourd, O. Toublan, J. Stevenson, and A. Walton, "Development of Electrical On-Mask CD Test Structures Based on Optical Metrology Features," in *Proceedings of*

*IEEE International Conference on Microelectronic Test Structures*, Tokyo, Japan, March 2007, pp. 171–176.

[6] ——, "Investigation of Electrical and Optical CD Measurement Techniques for the Characterisation of On-Mask GHOST Proximity Corrected Features," in *Proceedings of IEEE International Conference on Microelectronic Test Structures*, Edinburgh, UK, March 2008, pp. 29–34.

[7] S. Smith, A. Tsiamis, M. McCallum, A. Hourd, J. Stevenson, A. Walton, R. Dixson, R. Allen, J. Potzick, M. Cresswell, and N. Orji, "Comparison of Measurement Techniques for Advanced Photomask Metrology," in *Proceedings of IEEE International Conference on Microelectronic Test Structures*, Edinburgh, UK, March 2008, pp. 35–39.

[8] S. Smith, A. Tsiamis, M. McCallum, A. Hourd, J. Stevenson, and A. Walton, "Electrical Measurement of On-Mask Mismatch Resistor Structures," in *Proceedings of IEEE International Conference on Microelectronic Test Structures*, Tokyo, Japan, March 2007, pp. 3–8.

[9] ——, "On-Mask Mismatch Resistor Test Structures for the Characterisation of Maskmaking Capability," in *Proceedings of IEEE International Conference on Microelectronic Test Structures*, Edinburgh, UK, March 2008, pp. 228–232.

[10] T. Schaetz, S. Doebereiner, G. Scheuring, and H.-J. Brueck, "CD Metrology on OPC Features Using Light Optical and Electron Optical Tools," in *Proceedings of the SPIE, Vol. 4562: 21st Annual BACUS Symposium on Photomask Technology*, 2002, pp. 256–263.

[11] A. Tsiamis, S. Smith, M. McCallum, A. Hourd, J. Stevenson, and A. Walton, "Electrical Test Structures for the Characterisation of Optical Proximity Correction," in *Proceedings of the SPIE, Vol. 6533: The* $23^{rd}$ *European Mask and Lithography Conference (EMLC2007)*, Grenoble, France, January 2007, p. 65330M.

[12] M. Newsam, A. Walton, and M. Fallon, "Numerical Analysis of the Effect of Geometry on the Performance of the Greek Cross Structure," in *Proceedings of IEEE International Conference on Microelectronic Test Structures*, Trento, Italy, March 1996, pp. 247–252.

# Mapping the Edge Roughness of Test-Structure Features for Nanometer-Level CD Reference-Materials¶

M.W. Cresswell, M. Davidson,[1] G.I. Mijares, R.A. Allen, J. Geist, M. Bishop[2]

Semiconductor Electronics Division, Electronics and Electrical Engineering Laboratory
National Institute of Standards and Technology, Gaithersburg, Maryland, USA
[1] Spectel Research Corp., Palo Alto, California, USA
[2] International SEMATECH, Montopolis Drive, Austin, Texas, USA.

## ABSTRACT

The near-term objective of the work reported here is to develop a protocol for rapidly mapping CD and edge roughness from high-resolution SEM images of reference-material features patterned on Single-Crystal CD Reference Material (SCCDRM) chips. The longer term mission is to formulate a metric to enable automated characterization of as-fabricated reference-feature segments for rapid identification of fabrication-process enhancements and, ultimately, to select feature segments for further characterization as standard reference-materials. The selection of results presented here provides a new level of SCCDRM characterization which shows that segments of some SCCDRM features appear to have very useful extended lengths of up to 200 nm of superior CD uniformity.

## BACKGROUND

Fabrication processes and calibration procedures for making silicon-based prototype critical-dimension (CD) reference features available to industry have been under development in a multi-laboratory National Institute of Standards and Technology (NIST) project for several years.[1] The silicon technology has become known as the Single-Crystal CD Reference Material (SCCDRM) implementation. The end application for these reference materials is metrology support for sub 100 nm gate-length IC fabrication. The project has so far delivered a selection of reference materials for evaluation to the International SEMATECH Manufacturing Initiative and other organizations.[2] The better SCCDRM calibrated reference features with sub-tenth-micrometer linewidths that have so far been delivered have expanded uncertainties as low as 1.25 nm.‡ Navigation errors sustained during calibration team up with residual reference-feature CD non-uniformity to generate contributions of approximately 25 % of this amount. However, calibration with sub-single-nanometer uncertainty is understood to be highly desirable for end-user applications. The fact is that the significant uncertainty contribution above could be driven towards zero if a way could be found to fabricate reference features with CD roughness consistent with atomic-level feature-sidewall planarity along segment lengths as long as 0.25 μm, for example. Until very recently, the longest feature segments having this property extended for approximately 50 nm. However, further fabrication-process enhancement to generate greater lengths of quasi-atomic-sidewall-planarity is challenged by the possible existence of spatially random local regions of anomalous properties of the starting-material extending to over hundreds of nanometers. The approach that has been adopted here to minimize the adverse impact of these properties, and to produce 250 nm feature-segments having near-zero edge roughness, is to devise appropriate fabrication processing with the aid of a high-throughput data-acquisition protocol to identify regions of processing space that minimize, or zero, intra-feature CD and edge roughness. Simultaneously, the same data acquisition identifies the locations of segments on a feature having superior edge and CD roughness. The specific technical approach, namely automated dimensional analysis of high-resolution SEM (Scanning-Electron Microscopy) images, has been applied to a selection of SCCDRM features having CDs in the range 50 nm to 200 nm. The results provide an essential assessment of the current baseline SCCDRM fabrication process *vis-à-vis* CD-uniformity and edge roughness. This is a prerequisite to the next step of formulating a metric to enable automated identification and ranking of as-fabricated reference-feature segments to be used to support fabrication-process enhancements and, ultimately, to select features for further qualifying metrology for standard reference-material applications.[3]

One of many recent articles exemplifies how topical the issue of edge-roughness metrology has become.[4] However, there appear to be no prior reports on the unique and specific application that is the subject of this paper. Alternative methods include, for example,

¶ Contribution of the National Institute of Standards and Technology (NIST). Not subject to copyright.

‡ For a description of terminology see, for example, http://physics.nist.gov/cuu/Uncertainty/coverage.html .

performing the metrology on state-of-the-art production-level CD-SEM systems with built-in facilities for dimensional metrology. Unfortunately it is not practical to employ such systems for the preliminary laboratory research that is reported here because the approach that we have adopted is necessarily a chip-level implementation.

## PURPOSE

The mission of the work reported here has two parts. The first is to apply image-analysis software to extraction and analysis of multiple measurements of linewidth, linewidth roughness, and edge roughness of as-fabricated SCCDRM test-structure features from SEM images. The second part is to make an assessment of the status of the SCCDRM fabrication process for producing reference features with CD and edge roughness consistent with the longer term objective of fabricating Standard Reference Materials (SRMs) with sub-single-nanometer traceable uncertainty. The motivation for this study derives from the issue stated above: namely, that spatially random CD non-uniformity adversely impacts the level of uncertainty that is attributed to the certified CDs of deliverable reference features. The counter-measures to these problems that have so far been implemented include ranking and selecting features according to their linewidth uniformity and identifying the region of processing space that minimizes the observed non-uniformities. Since both are inspection-intensive operations, the specific longer-term target of the work reported here is providing a computer-based metric for fabrication-process enhancement and facilitating optimal selection of features to be delivered as SRMs for application in AFM (Atomic Force Microscopy), SEM, OCD (Optical CD), and/or SAXS (Small Angle X-Ray Scattering) environments. This report describes the analytical infrastructure that was developed and applied and presents a selection of the measurements that have so far been acquired. This sets the scene for future reports on the behavior of the dimensional parameter sets under various wafer-processing conditions.

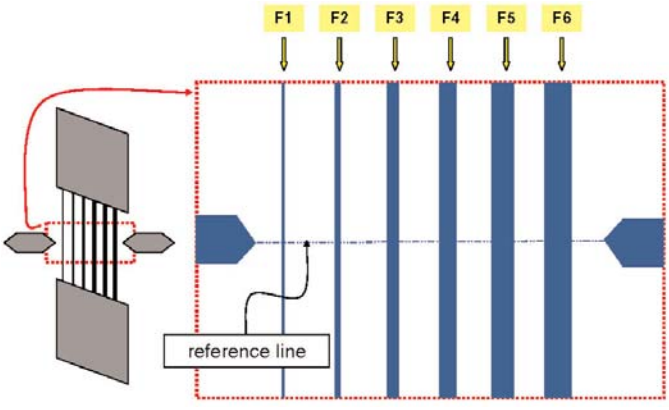



Figure 1. The labels F1 through F6 identify individual features of the test structure. F1 is the narrowest.

## TEST-CHIP DESIGN AND FABRICATION

### *Test-Chip and Test-Structure Design*

The test structure used in this project has multiple instances of a so-called "HRTEM (High-Resolution Transmission-Electron Microscopy) Target" that has six CD reference features. These six features were drawn with staggered linewidths so as to facilitate CD-extraction for multiple drawn-CD values. The basic geometry of an HRTEM target layout is shown in Figure 1. It was designed specifically to enable HRTEM imaging of the cross sections of all six features at a pre-determined location with a single sample-preparation. High-magnification top-down SEM images of all six features can be recorded in no more than two files. During lithography, the principal axes of the test structure are designed for alignment to <112> lattice vectors in the (110) surface of the SIMOX (Separation by Implantation with Oxygen) wafer to provide the pattern's features with planar (111) sidewalls.[5]

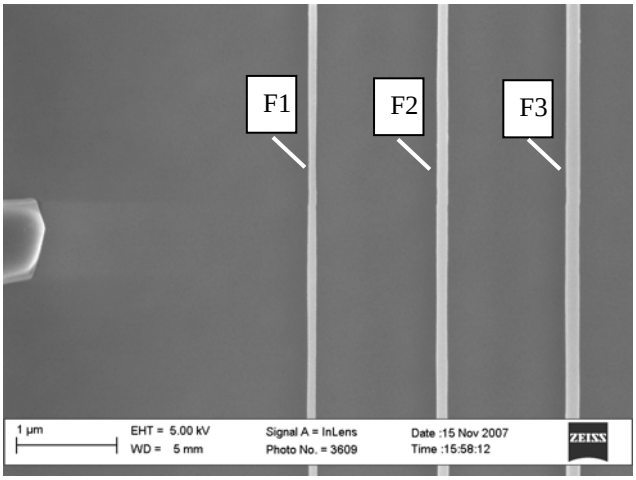



Figure 2. An SEM image of the left half of the central portion of the test structure shown in Figure 1.

### *Fabrication Process*

The test chip pattern is replicated in the device layer of a (110) SIMOX wafer according to the SCCDRM fabrication process.[5] The nominal heights of all the reference features are 150 nm. The device layer is electrically isolated from the remaining thickness of the substrate by the 390 nm-thick buried oxide created by high-energy oxygen implantation.

An SEM image of the left half of the as-fabricated central portion of the test structure that has been shown in Figure 1 is reproduced in Figure 2. One near term objective of the work reported here is to configure the Spectel Research "*MEASURE*" program to extract CD, CD-

roughness, left-edge roughness, and right-edge roughness parameters from multiple contiguous locations on all three features, of which those shown in Figure 2 are an example, in less than several seconds.[§] Whereas the as-distributed Spectel "*MEASURE*" program has a facility for automatically extracting the parameter values at one particular location on multiple images, the same is not true for multiple locations along a single feature on a single image, which is an essential component of the SCCDRM feature-ranking/selection task. Therefore the interim approach that was adopted here was to drive *MEASURE's* CD/roughness-extraction algorithm with a macro facility.

*The CD-Uncertainty Issue*

The complexities of the SCCDRM reference-feature calibration procedure have been described in detail elsewhere.[5] This synopsis explains why the issue of devising a convenient metric for fabrication-process enhancement is so important for further reduction of uncertainty.

The SCCDRM reference-feature program at NIST originated with AFM measurements of the linewidths of a selection of features that were subject, among others, to an error attributable to the unknown width of the AFM tip that was employed. A sub-selection of the measured features was then set aside for use internally for the future calibration of other AFM tips. The remaining features were then subjected to traceable HRTEM linewidth metrology. The AFM and HRTEM measurements were then reconciled to provide an estimate of the *AFM Offset* correction applicable to the features that were unavoidably destroyed during HRTEM imaging as well as those that were set aside. The set-aside features are thus able to serve as references for measuring the offsets applicable to replacement tips in future calibration of other reference materials.

The calibration of a newly fabricated reference material, with traceability to the HRTEM measurements described above, involves providing for uncertainty contributions from four sources.[5] The expanded uncertainty applicable to an AFM measurement of its CD is computed by combining the four contributions by "root-sum-squares" and multiplying the result, which is known as the "*combined standard uncertainty,*" by k=2 to generate the total 2-sigma uncertainty value. Among the four sources of the combined standard uncertainty that are referenced above, the one identified as *CD Non-Uniformity/Navigation* typically contributes approximately 0.2 nm or more to the total 2-sigma uncertainty. Currently, the better SCCDRM calibrated reference features have expanded uncertainties in excess of 1.25 nm. Thus the *CD Non-Uniformity/Navigation* contribution of 0.2 nm or more is not trivial in the ultimate quest to reduce the expanded uncertainty to below end-user-stated requirements of 1 nm. The fact is that this contribution could be driven to near zero if a way could be found to fabricate reference features with quasi-zero CD roughness along reference-feature lengths approaching 0.25 µm. Further reduction of uncertainty may also accrue through a second contributor to the expanded uncertainty, namely one labeled the "*Estimated AFM offset.*" Further details are provided in Reference 5.

The end-user of a SCCDRM reference material may also benefit from zero CD roughness while he or she navigates to the calibrated location of the reference feature. Residual roughness compounds the effect of spatially "missing" the calibrated location during, for example, a tip-width calibration.

The core reason that this work was undertaken is to reduce the zero CD-/edge-roughness condition to practice as far as possible, thereby reducing the calibrated-feature uncertainty. An essential prerequisite of the management of optimizing the fabrication process to provide at least local regions of features with zero edge roughness is having an extensive high-throughput data-acquisition protocol, an example of which is to be exemplified here, to assist in identifying optimum process conditions.

## TECHNICAL APPROACH

*SEM Imaging*

The individual feature-referencing designation that is used extensively in this document follows a format that has been used in prior articles on SCCDRM technology. The basic components of the designation uniquely identify each of typically 500 features that are replicated on each chip according to its :<Chip Number>, <Target-Array Number>, <Target Number>, and <Feature Number>. For example, A10-T3-5p3-F4 is a reference feature on chip A10, in the T3 HRTEM target array, specifically the 5p3 target, where it is the fourth feature. Examples of wafer and target-array maps are shown in Reference 5.

*The Spectel Research "MEASURE" Program*

The program *MEASURE* was developed with SEMATECH funding in the 1990s and is an offline metrology analysis tool designed for scanning electron microscope images. Besides providing a menu of conventional line-edge algorithms, with which to perform measurements, its capabilities were extended to include inverse scattering techniques based on Monte Carlo simulation. *MEASURE* can handle a wide variety of file

[§] Certain commercial equipment, instruments, or materials are identified in this document in order to specify adequate measurement procedures. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

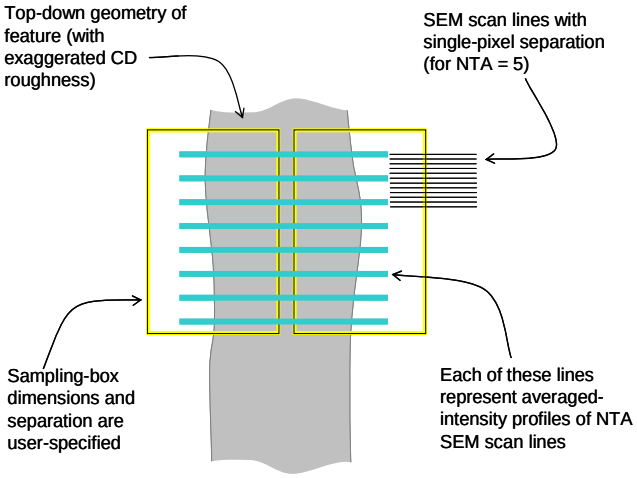



Figure 3. The intensity profiles of NTA rows (beam scans) are averaged before a user-selected line-edge finding algorithm is applied

formats, can perform automated measurements of folders containing many files, and includes some algorithm alignment capability. Its features can be further automated by using PC macro-writing tools such as *autoIT*, *autoHotKey*, or *MacroMaker*. A limited version of *MEASURE* is available free of charge from the Spectel Research FTP site.

When performing CD, CD-roughness, and line-edge roughness calculations, *MEASURE* uses a parameter called "The Number-of-Rows to Average" (NTA). The default for this parameter is 5 rows, each "row" being generated by a single scan of the beam, as depicted in Figure 3. The intensity profiles of this number of scans are averaged, and then a user-selected line-edge finding algorithm is applied to the averaged-intensity data. During measurement, each of a pair of sampling boxes with user-defined dimensions is located over the feature edges. The feature edges are found by a user-selected algorithm applied to the averaged profiles within the sampling boxes. The separation of the left and right edges is then calculated for each averaged set of rows. The CD for each sampling-box placement is then calculated as the average left-to-right separation. Similarly, for each placement, CD roughness is the standard uncertainty of the extracted CD value. The right- or left-edge roughness for the prevailing sampling box location is similarly calculated but by referencing the neighboring sampling-box boundaries.

*Measurement-Extraction Procedure*

For each of the reference features F1 thru F6 on each target, CD and roughness measurements were extracted from thirteen contiguous placements pairs of 300 nm long sampling boxes along the length of the imaged feature by the Spectel *MEASURE* program. This baseline protocol provided near-total coverage of their 3900 nm imaged lengths. For each feature, *MEASURE* thus provided 13 sets of 1 CD, 1 CD-roughness, and 2 edge roughness values. From this point on in this manuscript, these four quantities are referred to as the "dimensional parameter set" (DPS). The baseline NTA parameter was set at 5 rows per sampling-box.

The *MEASURE* program tabulates the extracted values of the dimensional parameter set in a "Report," an example of which is shown in Figure 4. In this case, it can be seen that the so-called "Maximum Derivative Algorithm" for DPS extraction was applied. It identifies the locations of the points where the derivative of pixel intensity vs. cross-section location is a maximum. It is one of several that *MEASURE* makes available to the user. For each sampling-box placement, the report lists the name of the image file, lists the local values of the dimensional parameter set, and identifies the actual algorithm that was selected to generate the latter. The summary page for each target lists the CDs of segments that report zero average CD roughness for all six features on a particular target.

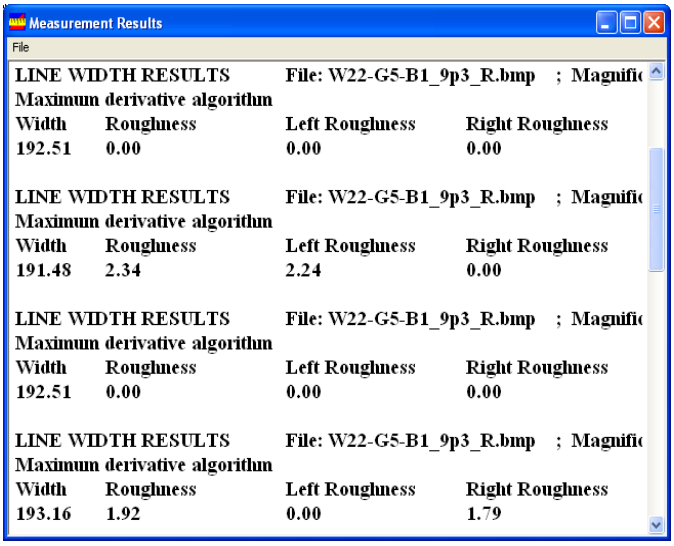



Figure 4. The Spectel *MEASURE* program tabulates the extracted values of the dimensional parameter set in a "Report." This Report is for four placements of the sampling boxes.

An open-source program called *MacroMaker* was used to drive *MEASURE* so as to compile a Report listing of 39 dimensional parameter sets acquired from each image of three features.[**]

*Examples of Variation of CD and Edge Roughness*

An example of a part of a worksheet listing of dimensional parameter set values for a six-feature HRTEM target is shown in Table 1. The yellow high

[**] http://www.sharewareplaza.com/MacroMaker-download_3600.html

| Calibration | Magnification = 200 KX SEM image | | | | |
|---|---|---|---|---|---|
| Box height | 300 | | Algorithm | Maximum derivative algorithm | |
| Recordings | F1, F2, & F3 | | Scan date | 12-14-08 | |
| Images | File: W22-G5-B1_9p3-200kx_L.bmp | | | File: W22-G5-B1_9p3_R.bmp | |

| Feature / Segment | F1 | F2 | F3 | F4 | F5 | F6 |
|---|---|---|---|---|---|---|
| S1 | | | 136.62 | | 192.51 | 211.14 |
| S2 | | | | 173.88 | 192.51 | |
| S3 | 86.94 | | | 173.88 | 192.51 | |
| S4 | | | | | | 217.35 |
| S5 | | | | 161.46 | | |
| S6 | | | 142.83 | 161.46 | 192.51 | |
| S7 | | 111.78 | | | | |
| S8 | | | | | | |
| S9 | | | | | | |
| S10 | | 105.57 | | | | |
| S11 | | | 142.83 | | 192.51 | |
| S12 | | | | | | |
| S13 | | | | | 192.51 | |

Figure 5. The summary page for each target reports the CDs, in nanometers, of segments that report zero average CD roughness for all six features on a particular target.

lighted cells indicate those segments for which the average CD roughness for a particular segment was reported as zero. Their distribution along the length of a particular feature is of central interest in the subject application because they identify segments that have greater potential usefulness as reference features and/or flag processing conditions that promote, or discourage, the patterning of segments having uniform CDs.

In addition, a summary page lists the highlighted CD data for all six features on a particular target. This allows a broader overview of the CD distribution across an entire target. An example of the summary page for the same target is shown in Figure 5, this time with the pixel counts in nanometers, as delivered by the *MEASURE* report.

*Examples of Variation of CD with Location*: During the initial phase of this work, data sets such as that shown in Table 1 were extracted from almost 100 features on three different chips. The few examples shown here have been selected by the authors on the basis of some preliminary observations; e.g., the curve in Figure 6 shows a degree of full-length uniformity, other than for a sharp drop in CD around the 2100 nm mark. The left-most section that is comprised of segments S1 thru S7 appears to have a CD-uniformity of better than 1 nm over 2.0 µm. On the face of it, this region would be a candidate for inspection for a reference-material application.

Another observation, which has been made on several CD profiles, is also exemplified by Table 1 and Figure 6. Namely, contiguous segment-sets having zero CD roughness often have among the highest CDs along the feature lengths. What is somewhat surprising is that generally, but not always, the segments of a particular feature having zero CD roughness also have the *same* segment-average CDs. Both of these points are exemplified by the data in Table 1 and Figure 6.

*Example of Edge-Roughness Results:* In this and the following sections, the examples are drawn largely from images collected from target G5-B1-9p3 because it usefully illustrates some of the points from which provisional observations may be made on the status and characteristics of SCCDRM fabrication.

Table 1. Example of a part of a feature-page listing of Dimensional-Parameter-Set values generated from a Report such as that shown in Figure 4. The feature is W22-G5-B1-9p3-F5, and the units are pixels, each pixel calibrated as 6.21 nm.

| Average segment CD | Average CD Roughness | Left-edge CD Roughness | Right-edge CD Roughness | Segment |
|---|---|---|---|---|
| 31.00 | 0.00 | 0.00 | 0.00 | S 1 |
| 31.00 | 0.00 | 0.00 | 0.00 | S 2 |
| 31.00 | 0.00 | 0.00 | 0.00 | S 3 |
| 30.94 | 0.24 | 0.22 | 0.00 | S 4 |
| 30.96 | 0.20 | 0.19 | 0.00 | S 5 |
| 31.00 | 0.00 | 0.00 | 0.00 | S 6 |
| 31.10 | 0.31 | 0.00 | 0.28 | S 7 |
| 30.10 | 0.31 | 0.00 | 0.26 | S 8 |
| 30.27 | 0.45 | 0.00 | 0.43 | S 9 |
| 30.44 | 0.50 | 0.00 | 0.25 | S 10 |
| 31.00 | 0.00 | 0.00 | 0.00 | S 11 |
| 30.98 | 0.14 | 0.14 | 0.00 | S 12 |
| 31.00 | 0.00 | 0.00 | 0.00 | S 13 |

In the fabrication of features with sub 50 nm linewidths, our experience has been that isolated features, even when replicated by optical ultra-violet lithography, can have their linewidths driven down to this range by prolonging the etching with which they are patterned. The limiting factor, as far as the reference-material application is

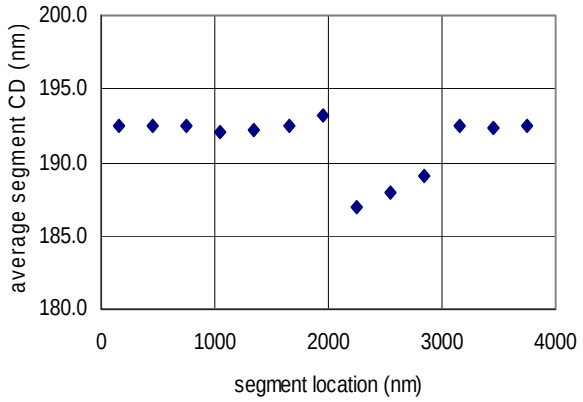



Figure 6. CD averaged over 300 nm vs. segment location on the feature. The zero location is the extreme top of the image and the 4000 nm location is at the bottom, consistent with the orientation shown in Figure 1 and Figure 3.

concerned, is replicating features with superior CD uniformity. Experience in the inspection and rating as-produced features with the SEM-MEASURE approach reports results for CD-variation patterns quite similar to those generated by the AFM approach. It happens that the former is much more rapid and, therefore, can be applied much more extensively. Of course, it is not traceability-capable while ranking features for superior uniformity, as is the AFM approach, which is uniquely responsive to the traceability issue.

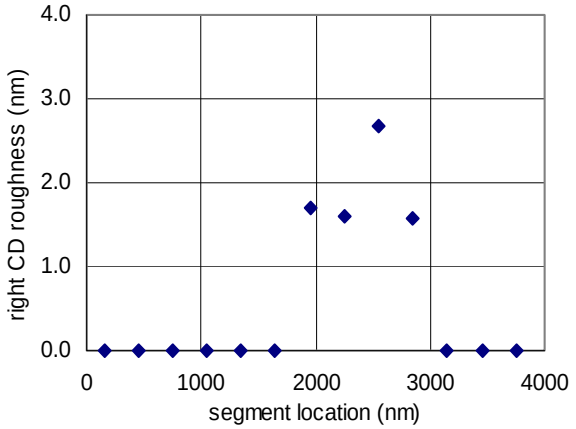



Figure 7. Non-zero right-edge roughness occurs almost exactly along the length of the feature where the local CD degradation shown previously in Figure 5 occurs. (Feature is W22-G5-B1-9p3-F5).

The right-edge roughness profile shown in Figure 7 also derives from Table 2 and illustrates that anomalous right-edge roughness occurs almost exactly along the length of the feature where the local CD degradation that has been shown previously in Figure 6 occurs. Whereas we later propose a physical model for how this could happen, such behavior has not yet been widely cataloged.

So far in this paper, all the reported measurements were conducted with the *MEASURE's* "Maximum Derivative" (MDA) algorithm for linewidth extraction. We also examined the repeatability the G5-B1-9p3-F5 feature measurements with *MEASURE's* alternative "Maximum Second Derivative" (MSD) algorithm. The repeatability *per se* was somewhat improved, and both the linewidth and roughness values closely tracked those generated by the MDA algorithm. If there was a repeatability issue, our preliminary observations indicate that the application of MSD might be preferable.

## SUMMARY

The results that have been presented here show that segments of some SCCDRM features appear to have very useful extended lengths of up to 200 nm of superior CD uniformity. In general, however, these segments have no predictable spatial distribution. A major unintended benefit of the results that have been generated is that we may have acquired some useful insight for understanding the etching process that is used to pattern the features. We have observed that (a) the most uniform feature segments tend to have the highest local CDs along a feature length and (b) such CDs tend to have the same values. As a result, a model for the generation of the non-uniformity of the CD, which in principle might be expected to be quasi-atomically uniform, is that the starting material may have nanometer-scale regions, which, for an as-yet unknown reason, destroy local etch anisotropy and accelerate silicon dissolution.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr Craig D. McGray of the Semiconductor Electronics Division and William F. Guthrie of the Statistical Engineering Division at NIST for reviewing the manuscript. Stefano Nungovich, who was a 2008 Summer Intern in the NIST Semiconductor Electronics Division, rendered early assistance in evaluating some of the software that was used. The authors benefited extensively from their earlier collaboration with Dr. R. G. Dixson of the NIST Precision Engineering Division. Their results were developed from extensive collaboration in wafer-processing development with Mr. Rene Blanquies and Dr. Marco Tortonese of VLSI Standards, Inc.

## REFERENCES

1. R. A. Allen, B. A. am Ende, M. W. Cresswell, C. E. Murabito, T. J. Headley, W. F. Guthrie, L. W. Linholm, C. H. Ellenwood, and E. H. Bogardus, Test Structures for Referencing Electrical Line-width Measurements to Silicon Lattice Parameters Using HRTEM, IEEE Transact. Semiconductor Manufacturing, 16 (2), pp. 239-248 (2003).

2. R. G. Dixson, R. A. Allen, W. F. Guthrie, and M. W. Cresswell, Traceable Calibration of Critical-Dimension Atomic Force Microscope Linewidth Measurements with Nanometer Uncertainty, J. Vac. Sci. Technol. B Volume 23, Issue 6, pp. 3028-3032 (November 2005).

3. R. A. Allen, T. J. Headley, S. C. Everist, R. N. Ghoshtagore, M. W. Cresswell, and L. W. Linholm, "High-Resolution Transmission Electron Microscopy Calibration of Critical Dimension (CD) Reference Materials," *IEEE Transactions on Semiconductor Manufacturing*, Vol. 14, No. 1, pp. 26–31 (2001).

4. N. G Orji, T. V Vorburger, J. Fu1, R. G Dixson, C. V. Nguyen, and J. Raja, "Line Edge Roughness Metrology Using Atomic Force Microscopes," Meas. Sci. Technol. 16 (2005).

5. M. W. Cresswell,; W. F. Guthrie,; R. G. Dixson,; C. E. Murabito,; J. V. M.De Pinillos, J. Res. Natl. Inst. Stand. Technol., 111, 187 (2006).

# $S$-Parameter-Based Modal Decomposition of Multiconductor Transmission Lines and Its Application to De-Embedding

S. Amakawa[1], K. Yamanaga[1], H. Ito[2], T. Sato[1], N. Ishihara[1], and K. Masu[1,2]

[1]Integrated Research Institute, Tokyo Institute of Technology
[2]Precision and Intelligence Laboratory, Tokyo Institute of Technology

## ABSTRACT

Theory and experiments are presented of modal decomposition of scattering matrices of multiconductor transmission lines (TLs). In effect, $n$ coupled TLs are decomposed into $n$ independent ones. Its use is demonstrated by applying it to thru-only de-embedding of 4 coupled TLs (synthesized data) and 2 coupled TLs (measurement data from a 0.18 $\mu$m-CMOS chip). The proposed de-embedding method could greatly facilitate accurate characterization of on-chip multiport networks.

## INTRODUCTION

Multiconductor transmission lines (MTLs) have been a subject of intensive study for decades, and modal analysis has been a very important tool [1]–[4]. MTL equations are typically written in terms of per-unit-length equivalent-circuit parameters. Experimental characterization of MTLs, therefore, often involves extraction of those parameters from measurement data [5], [6], which often being scattering matrices or $S$ matrices. When the device under test (DUT) is on-chip or on-board TLs, as-measured $S$ matrices usually include properties of intervening structures such as probe pads. Then, extraction of equivalent-circuit parameters has to be done after de-embedding [7] the intrinsic properties of the TLs.

Among a number of de-embedding methods, thru-only methods [8]–[13] are particular attractive because de-embedding procedures may be formulated in such a way that no particular assumption is made about the physical topology of each half of the embedding network. A simple thru-only de-embedding method has recently been applied successfully to characterization of on-chip transmission lines (2-ports) up to a millimeter-wave range [10], [13]. The method was subsequently extended to 4-port networks and applied to de-embedding of differential TLs [14]. The essential idea in the latter [14] was to reduce a 4-port problem to two independent 2-port problems by a coordinate transformation. The requirement for it to work was that the $4\times4$ $S$ matrix of the THRU dummy pattern (a pair of nonuniform TLs) be block-diagonalizable by the even/odd transformation, meaning that the THRU has a certain symmetry property. This development naturally leads to the idea that the same de-embedding method should be applicable to $2n$-ports, where $n$ is a positive integer, provided that the $S$ matrix of the THRU ($n$ coupled nonuniform TLs) can somehow be block-diagonalized with $2\times2$ diagonal blocks. In this paper, we show how the desired transformation can be performed and demonstrate its use by thru-only de-embedding of modeled 4 coupled TLs (an 8-port) and measured 2 coupled TLs (a 4-port). We assume throughout that the THRU patterns under measurement are reciprocal and hence the associated $S$ matrices symmetric.

## DECOMPOSITION OF A $2n$-PORT INTO $n$ 2-PORTS

Our goal here is to transform a $2n\times2n$ scattering matrix $\boldsymbol{S}$ into the following block-diagonal form:

$$\tilde{\boldsymbol{S}}' = \begin{bmatrix} \boldsymbol{S}_{\mathrm{m}1} & & \\ & \ddots & \\ & & \boldsymbol{S}_{\mathrm{m}n} \end{bmatrix}, \tag{1}$$

where $\boldsymbol{S}_{\mathrm{m}i}$ are $2\times2$ submatrices and the rest of the elements of $\tilde{\boldsymbol{S}}'$ are all 0. The port numbering scheme for $\tilde{\boldsymbol{S}}'$ is shown in Fig. 1 with primes. Once the transformation is done, the DUT can be treated as if they were composed of $n$ uncoupled 2-ports. This is not an ordinary matrix diagonalization problem. The form of (1), however, results by the artifice of reordering the rows and columns of $\tilde{\boldsymbol{S}}$ in (2), consisting of four diagonal submatrices, such that $\boldsymbol{S}_{\mathrm{m}i}$ in (1) is built from the $i$th diagonal elements of the four submatrices of $\tilde{\boldsymbol{S}}$.

$$\tilde{\boldsymbol{S}} = \left[\begin{array}{c|c} \ddots & \ddots \\ \hline \ddots & \ddots \end{array}\right]. \tag{2}$$

The port indices of $\tilde{\boldsymbol{S}}$ are shown in Fig. 1 without primes. The problem, therefore, is the transformation of $\boldsymbol{S}$ into $\tilde{\boldsymbol{S}}$ followed by the reordering of rows and columns yielding $\tilde{\boldsymbol{S}}'$.

A generalized scattering matrix $\boldsymbol{S}$ of a $2n$-port relates the vector, $\boldsymbol{a}$, of power waves of a given frequency incident on the $2n$-port to the vector, $\boldsymbol{b}$, of outgoing power waves [15]:

$$\boldsymbol{b} = \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{bmatrix} = \boldsymbol{S}\boldsymbol{a}. \tag{3}$$

In the case of a cascadable $2n$-port, it makes sense to divide the ports into two groups as shown in Fig. 1, and hence the division of $\boldsymbol{S}$, $\boldsymbol{a}$, and $\boldsymbol{b}$ into submatrices/subvectors in (3). If the $2n$-port is reciprocal, $\boldsymbol{S}$ is symmetric: $\boldsymbol{S}^{\mathrm{T}} = \boldsymbol{S}$, where $^{\mathrm{T}}$ denotes transpose. Then, it can be shown that the following change of bases brings about the desired transformation.

$$\begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{W}_1 & \\ & (\boldsymbol{W}_2^{\mathrm{T}})^{-1} \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{a}}_1 \\ \tilde{\boldsymbol{a}}_2 \end{bmatrix}, \tag{4}$$

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

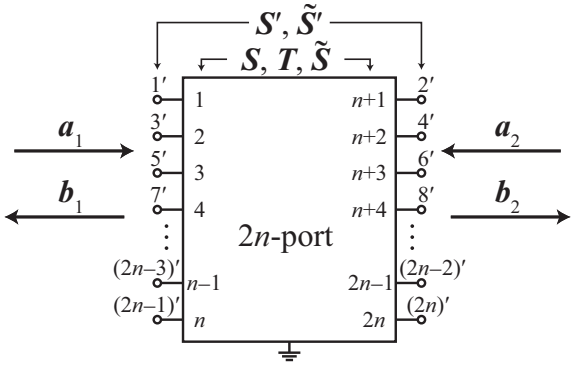



Fig. 1. Port indices for a cascadable $2n$-port. The ports 1 through $n$ of $\boldsymbol{S}$ constitute one end of the bundle of $n$ lines and the ports $n+1$ through $2n$ the other end. The indices with a prime are useful after the modal decomposition into 2-ports.

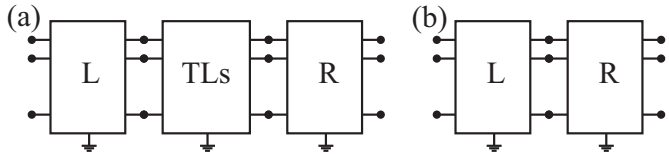



Fig. 2. (a) A model of $n$ coupled TLs measured by a VNA. The TLs sit between the intervening structures L and R. (b) A model of a THRU.

$$\begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix} = \begin{bmatrix} (\boldsymbol{W}_1^{\mathrm{T}})^{-1} & \\ & \boldsymbol{W}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{b}_1} \\ \tilde{\boldsymbol{b}_2} \end{bmatrix}, \tag{5}$$

where the blanks represent zero submatrices. $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ diagonalize $\boldsymbol{S}_{21}^{-1}\boldsymbol{S}_{22}\boldsymbol{S}_{12}^{-1}\boldsymbol{S}_{11}$ and $\boldsymbol{S}_{22}\boldsymbol{S}_{12}^{-1}\boldsymbol{S}_{11}\boldsymbol{S}_{21}^{-1}$, respectively, by similarity transformation and can be computed by eigenvalue decomposition. The derivation parallels that presented in [4]. $\tilde{\boldsymbol{S}}$ is thus given by

$$\tilde{\boldsymbol{S}} = \begin{bmatrix} \boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{S}_{11}\boldsymbol{W}_1 & \boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{S}_{12}(\boldsymbol{W}_2^{\mathrm{T}})^{-1} \\ \boldsymbol{W}_2^{-1}\boldsymbol{S}_{21}\boldsymbol{W}_1 & \boldsymbol{W}_2^{-1}\boldsymbol{S}_{22}(\boldsymbol{W}_2^{\mathrm{T}})^{-1} \end{bmatrix}. \tag{6}$$

## DE-EMBEDDING USING A THRU

Thru-only de-embedding methods [8]–[13] assume that the device under measurement can be represented as shown in Fig. 2(a), in which the TLs to be characterized are embedded in parasitic elements such as pads. In terms of the transfer matrix $\boldsymbol{T}$ defined by

$$\begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{b}_1 \end{bmatrix} = \boldsymbol{T} \begin{bmatrix} \boldsymbol{b}_2 \\ \boldsymbol{a}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_2 \\ \boldsymbol{a}_2 \end{bmatrix}, \tag{7}$$

$$\begin{bmatrix} \boldsymbol{T}_{11} & \boldsymbol{T}_{12} \\ \boldsymbol{T}_{21} & \boldsymbol{T}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}_{21}^{-1} & -\boldsymbol{S}_{21}^{-1}\boldsymbol{S}_{22} \\ \boldsymbol{S}_{11}\boldsymbol{S}_{21}^{-1} & \boldsymbol{S}_{12} - \boldsymbol{S}_{11}\boldsymbol{S}_{21}^{-1}\boldsymbol{S}_{22} \end{bmatrix}, \tag{8}$$

$$\begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}_{21}\boldsymbol{T}_{11}^{-1} & \boldsymbol{T}_{22} - \boldsymbol{T}_{21}\boldsymbol{T}_{11}^{-1}\boldsymbol{T}_{12} \\ \boldsymbol{T}_{11}^{-1} & -\boldsymbol{T}_{11}^{-1}\boldsymbol{T}_{12} \end{bmatrix}, \tag{9}$$

the overall, as-measured $T$ matrix for Fig. 2(a) is $\boldsymbol{T}_{\mathrm{meas}} = \boldsymbol{T}_{\mathrm{L}}\boldsymbol{T}_{\mathrm{TL}}\boldsymbol{T}_{\mathrm{R}}$.

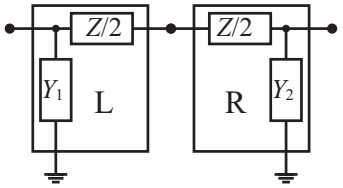



Fig. 3. A $\pi$-equivalent for representing a decomposed THRU in the modal domain.

In order to de-embed $\boldsymbol{T}_{\mathrm{TL}}$ from $\boldsymbol{T}_{\mathrm{meas}}$, a THRU pattern (Fig. 2(b)) is measured. The result ($\boldsymbol{T}_{\mathrm{thru}} = \boldsymbol{T}_{\mathrm{L}}\boldsymbol{T}_{\mathrm{R}}$) is transformed into the block-diagonal form $\tilde{\boldsymbol{S}}'_{\mathrm{thru}}$. Since each of the resultant $2 \times 2$ diagonal blocks of $\tilde{\boldsymbol{S}}'_{\mathrm{thru}}$ is symmetric by assumption, each modal 2-port can be represented by, for instance, a $\pi$-equivalent (Fig. 3). Note that the use of other representations having the same degree of freedom (e.g. a T-equivalent) is equally valid. The remaining task is to split the $\pi$-equivalent into two. One straightforward way of doing it is to split the series impedance $Z$ in half as shown in Fig. 3. The physical structure of a THRU can often be designed so that this splitting is reasonable. Since each half of Fig. 3 has sufficiently small number of parameters, the corresponding $S$ matrices, $\tilde{\boldsymbol{S}}_{\mathrm{L}}$ and $\tilde{\boldsymbol{S}}_{\mathrm{R}}$, can be determined as will be explained shortly. $\boldsymbol{T}_{\mathrm{L}}$ and $\boldsymbol{T}_{\mathrm{R}}$ are found by transforming $\tilde{\boldsymbol{S}}_{\mathrm{L}}$ and $\tilde{\boldsymbol{S}}_{\mathrm{R}}$ back to the original conductor domain and then converting them into $T$ matrices. Then, the characteristics of the TLs are obtained by $\boldsymbol{T}_{\mathrm{TL}} = \boldsymbol{T}_{\mathrm{L}}^{-1}\boldsymbol{T}_{\mathrm{meas}}\boldsymbol{T}_{\mathrm{R}}^{-1}$.

The determination of the parameters in Fig. 3 involves conversion of the modal 2-port $S$ matrices into admittance matrices through the general formula (˜ to be added when used in the modal domain)

$$\boldsymbol{Y} = \left(\boldsymbol{R}_0^{-1/2}\boldsymbol{Z}_0^{*} + \boldsymbol{S}\boldsymbol{R}_0^{-1/2}\boldsymbol{Z}_0\right)^{-1}(\boldsymbol{1}_{2n} - \boldsymbol{S})\,\boldsymbol{R}_0^{-1/2}, \tag{10}$$

and this requires an appropriate reference impedance matrix $\boldsymbol{Z}_0$. Here $^*$ denotes complex conjugate and $\boldsymbol{R}_0$ is the real part of $\boldsymbol{Z}_0$, namely $\boldsymbol{R}_0 = \Re(\boldsymbol{Z}_0)$. $\boldsymbol{Z}_0$ is usually a diagonal matrix in the conductor domain. When $\boldsymbol{S}$ is transformed into $\tilde{\boldsymbol{S}}$, $\boldsymbol{Z}_0$ also undergoes a transformation into $\tilde{\boldsymbol{Z}}_0$ [14]. One potential problem here is that, depending on the transformation matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, $\tilde{\boldsymbol{Z}}_0$ might become nondiagonal. This makes it difficult to associate the conductor-domain voltages and currents with their modal-domain counterparts. However, since our purpose of performing the transformation here is to facilitate de-embedding, we can actually use a *fictitious* reference impedance matrix of our choice in (10). A practitioner's choice might be $\tilde{\boldsymbol{Z}}_0 = 50 \cdot \boldsymbol{1}_{2n}$, where $\boldsymbol{1}_{2n}$ is a $2n \times 2n$ identity matrix. Then, the transformed network matrix will represent a physically different, fictitious system. But when transformed back to the conductor domain, no problem arises.

Let the admittance matrix of the $i$th modal 2-port obtained

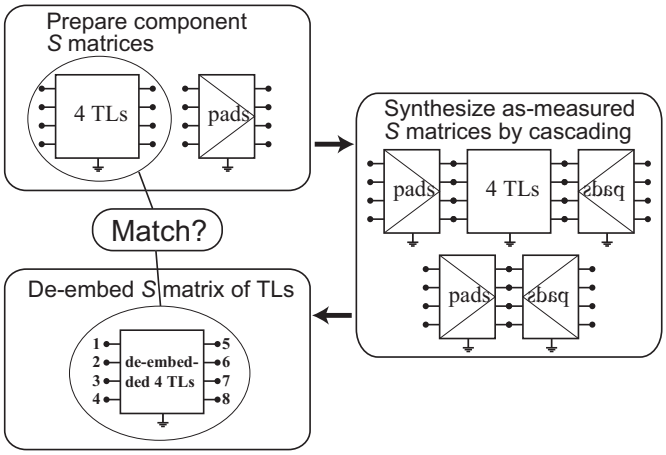



Fig. 4. The flow of validating the de-embedding method.

from (10) be

$$\tilde{\boldsymbol{Y}}_{\mathrm{thru},i} = \begin{bmatrix} y_{11} & y_{21} \\ y_{21} & y_{22} \end{bmatrix}. \tag{11}$$

Note that (11) is symmetric by the reciprocity assumption. The parameters in Fig. 3 are given by

$$Y_1 = y_{11} + y_{21}, \tag{12}$$

$$Y_2 = y_{22} + y_{21}, \tag{13}$$

$$Z = -1/y_{21}. \tag{14}$$

Note that it is also possible to determine the THRU using some other method (e.g. [11]).

## VALIDATION OF THE DE-EMBEDDING METHOD

The procedure that we followed for validating the thru-only de-embedding is shown in Fig. 4. $S$-parameter files of 1 mm-long 4 coupled TLs and pads were generated by using Agilent Technologies ADS. A cross section of the TLs is shown in Fig. 5. The schematic diagram representing the pads placed at each end of the bundle of TLs is shown in Fig. 6. Figs. 7 and 8 show the characteristics of the "as-measured" TLs and the THRU, respectively. The characteristics of the bare TLs and the de-embedded results are both shown on the same Smith chart in Fig. 9, but they are indistinguishable, thereby demonstrating the validity of the de-embedding procedure.

We also applied the proposed de-embedding method to a pair of TLs in 0.18 $\mu$m-CMOS (Fig. 10) previously analyzed by the even/odd transformation in [14]. The frequency ranged from 100 MHz to 50 GHz. The numerical values of $\boldsymbol{S}'$, $\boldsymbol{S}'_{\mathrm{e/o}}$ ($S$ matrix in the even/odd domain), and $\tilde{\boldsymbol{S}}'$ for the THRU (Fig. 10(a)) at 10 GHz are, respectively,

$$\begin{bmatrix} 0.011-0.094\mathrm{j} & 0.866-0.345\mathrm{j} & 0.040+0.030\mathrm{j} & -0.039-0.022\mathrm{j} \\ 0.866-0.345\mathrm{j} & 0.011-0.093\mathrm{j} & -0.038-0.024\mathrm{j} & 0.041+0.030\mathrm{j} \\ 0.040+0.030\mathrm{j} & -0.038-0.024\mathrm{j} & 0.009-0.093\mathrm{j} & 0.865-0.346\mathrm{j} \\ -0.039-0.022\mathrm{j} & 0.041+0.030\mathrm{j} & 0.865-0.346\mathrm{j} & 0.010-0.094\mathrm{j} \end{bmatrix},$$

$$\begin{bmatrix} 0.050-0.064\mathrm{j} & 0.828-0.369\mathrm{j} & 0.001-0.000\mathrm{j} & 0.001-0.000\mathrm{j} \\ 0.828-0.369\mathrm{j} & 0.051-0.064\mathrm{j} & -0.000+0.002\mathrm{j} & 0.001+0.000\mathrm{j} \\ 0.001-0.000\mathrm{j} & -0.000+0.002\mathrm{j} & -0.030-0.124\mathrm{j} & 0.904-0.322\mathrm{j} \\ 0.001-0.000\mathrm{j} & 0.001+0.000\mathrm{j} & 0.904-0.322\mathrm{j} & -0.030-0.123\mathrm{j} \end{bmatrix},$$

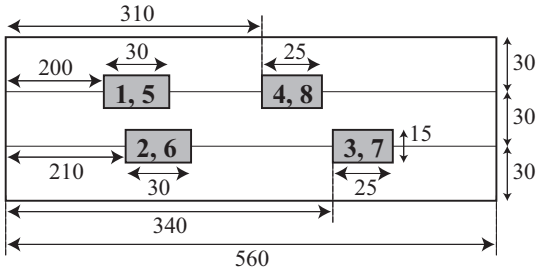



Fig. 5. A schematic cross section of the 1 mm-long 4 coupled TLs (not to scale), labeled with port numbers. Dimensions are in $\mu$m. Relative dielectric permittivity is 4. Metal conductivity is $5.9 \times 10^7\ (\Omega \cdot \mathrm{m})^{-1}$. $\tan\delta = 0.04$.

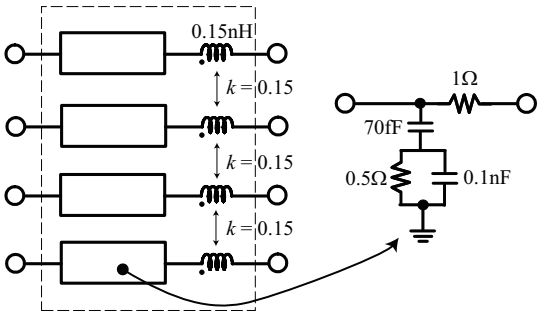



Fig. 6. A model of the left half of a THRU including pads.

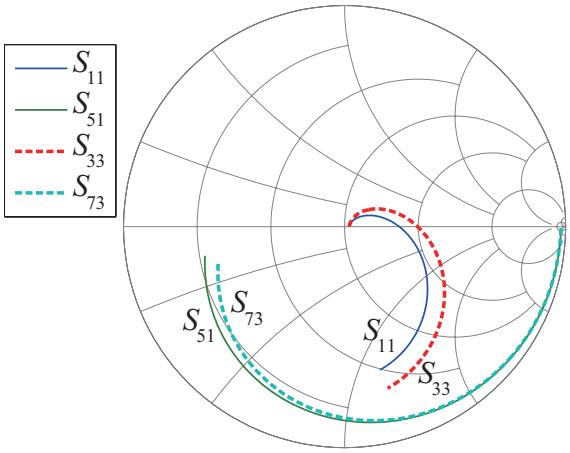



Fig. 7. Characteristics of the 4 coupled TLs from 100 MHz to 40 GHz before de-embedding.

$$\begin{bmatrix} 0.051-0.064\mathrm{j} & 0.828-0.369\mathrm{j} & 0.000+0.000\mathrm{j} & 0.000+0.000\mathrm{j} \\ 0.828-0.369\mathrm{j} & 0.050-0.064\mathrm{j} & 0.000+0.000\mathrm{j} & 0.000+0.000\mathrm{j} \\ 0.000+0.000\mathrm{j} & 0.000+0.000\mathrm{j} & -0.030-0.124\mathrm{j} & 0.903-0.324\mathrm{j} \\ 0.000+0.000\mathrm{j} & 0.000+0.000\mathrm{j} & 0.903-0.324\mathrm{j} & -0.032-0.123\mathrm{j} \end{bmatrix}.$$

Their corresponding reference impedance matrices are $\boldsymbol{Z}'_0 = \boldsymbol{Z}'_{0\mathrm{e/o}} = \tilde{\boldsymbol{Z}}'_0 = 50 \cdot \mathbf{1}_4$. The upper diagonal block in $\boldsymbol{S}'_{\mathrm{e/o}}$ is the even-mode $S$ matrix and the lower diagonal block is the odd-mode $S$ matrix. The residual nonzero off-diagonal blocks in $\boldsymbol{S}'_{\mathrm{e/o}}$, representing the crosstalk between the two modes, were ignored in [14]. As seen above, our proposed method can better block-diagonalize the $S$ matrix.

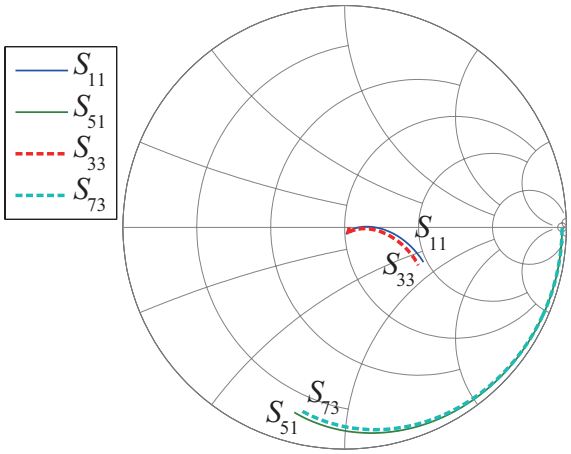



Fig. 8. Characteristics of the THRU from 100 MHz to 40 GHz.

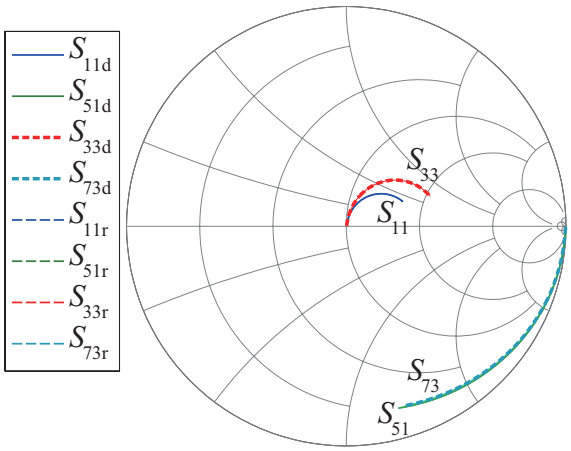



Fig. 9. Reference characteristics of the 4 coupled TLs (with a subscript r) and the de-embedded results. (with a subscript d). Actually, those two are indistinguishable on the Smith chart.

## CONCLUSIONS

We have proposed $S$-parameter-based modal decomposition of multiconductor transmission lines (MTLs) and its application to de-embedding. It has been established that the proposed method enables one to decompose an $S$ matrix of $n$ coupled TLs into that of $n$ uncoupled 2-ports. Then, a thru-only de-embedding method can be applied, provided that each half of the THRU can be determined in the modal domain. The proposed de-embedding method could be useful for accurate frequency-domain characterization of multiport networks such as transistors and bus lines consisting of multiple differential TLs.

## ACKNOWLEDGMENTS

This work was partially supported by JSPS.KAKENHI, MEXT.KAKENHI, MIC.SCOPE, STARC, and VDEC in collaboration with Agilent Technologies Japan, Ltd., Cadence Design Systems, Inc., and Mentor Graphics, Inc.

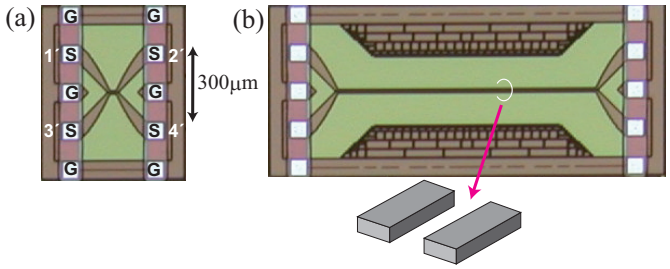



Fig. 10. (a) A micrograph of the THRU with GSGSG pads. (b) A pair of 1 mm-long TLs with the same pads. The line width is 6 $\mu$m and the spacing between the lines is 4.6 $\mu$m. The technology is a 0.18-$\mu$m-CMOS.

## REFERENCES

[1] C. R. Paul, *Analysis of Multiconductor Transmission Lines*, 2nd ed., Wiley-Interscience, 2008.

[2] H. Kogo, "A study of multielement transmission lines," IRE Trans. Microw. Theory Tech., vol.8, no.2, pp.136–142, March 1960.

[3] D. F. Williams, L. A. Hayden, and R. B. Marks, "A complete multimode equivalent-circuit theory for electrical design," J. Res. Natl. Inst. Stand. Technol., vol.102, no.4, pp.405–423, July–August 1997.

[4] J. A. B. Faria, "A new generalized modal analysis theory for nonuniform multiconductor transmission lines," IEEE Trans. Power Syst., vol.19, no.2, pp.926–933, May 2004.

[5] J. van der Merwe, H. C. Reader, and J. H. Cloete, "$S$-parameter measurements yielding the characteristic matrices of multiconductor transmission lines," IEEE Trans. Electromagn. Compat., vol.40, no.3, pp.249–256, August 1998.

[6] J. G. Nickel, D. Trainor, and J. E. Schutt-Ainé, "Frequency-domain-coupled microstrip-line normal-mode parameter extraction from $S$-parameters," IEEE Trans. Electromagn. Compat., vol.43, no.4, pp.495–503, November 2001.

[7] R. F. Bauer and P. Penfield, Jr, "De-embedding and unterminating," IEEE Trans. Microw. Theory Tech., vol.22, no.3, pp.282–288, March 1974.

[8] T. E. Kolding, "On-wafer calibration techniques for giga-Hertz CMOS measurements," Int. Conf. Microelectronic Test Struct., pp.105–110, 1999.

[9] D. C. Laney, *Modulation, Coding and RF Components for Ultra-Wideband Impulse Radio*, PhD thesis, University of California, San Diego, San Diego, California, 2003.

[10] Y. Tretiakov, K. Vaed, D. Ahlgren, J. Rascoe, S. Venkatadri, and W. Woods, "On-wafer de-embedding techniques for SiGe/BiCMOS/RFCMOS transmission line interconnect characterization," Proc. Int. Interconnect Technol. Conf., pp.166–168, 2004.

[11] E. S. Daniel, N. E. Harff, V. Sokolov, S. M. Schreiber, and B. K. Gilbert. "Network analyzer measurement de-embedding utilizing a distributed transmission matrix bisection of a single THRU structure," 63rd Automatic RF Techniques Group (ARFTG) Conf., pp. 61–68, June 2004

[12] Y. Goto, Y. Natsukari, and M. Fujishima, "New on-chip de-embedding for accurate evaluation of symmetric devices," Jpn. J. Appl. Phys., vol.47, no.4, pp.2812–2816, April 2008.

[13] H. Ito and K. Masu, "A simple through-only de-embedding method for on-wafer S-parameter measurements up to 110 GHz," IEEE MTT-S International Microwave Symposium Digest, pp.383–386, June 2008.

[14] S. Amakawa, H. Ito, N. Ishihara, and K. Masu, "A simple de-embedding method for characterization of on-chip four-port networks," Advanced Metallization Conf., pp.105–106, September 2008; Proc. Advanced Metallization Conf. 2008 (to be published).

[15] K. Kurokawa, "Power waves and the scattering matrix," IEEE Trans. Microw. Theory Tech., vol.13, no.2, pp.194–202, March 1965.

# Characterization and Model Parameter Extraction of Symmetrical Centre Tapped Inductor using Build in Mixed Mode and Pure Differential S-Parameters

F. Gianesello, Y. Morandini, S. Boret and D. Gloria
STMicroelectronics, Technology R&D, 850 avenue Jean Monnet, 38926 Crolles
Email: frederic.gianesello@st.com, Phone number: (033) 0476925541, Fax: (033) 0476925732

## I: Introduction

With the continuous reduction of the gate length, the cut off frequency of the active devices in CMOS technology has exceeded 200 GHz [1]. In addition, CMOS possesses the capability to integrate transceiver with the baseband circuits. Thus, CMOS technology seems to be an attractive candidate for low-gigahertz (5 GHz) radio frequency (RF) applications [2] and even millimetre wave one [3]. Then, monolithic inductors have become an important component in highly integrated radio frequency circuits (RF ICs) for wireless communication systems such as personal communication services, wireless local area networks, satellite communications, and the global positioning system. It is well known that exciting a spiral inductor differentially, using a source connected between the two ends of the inductor, the peak Q-factor shows a significant increase, and this high Q value is maintained over a broader bandwidth [4] compared to single-ended excitation. Up to now, differential characterization of symmetrical inductor has been performed using classical single ended S-parameters [5], and assuming linearity and electrical symmetry hypothesis, mixed mode S-parameters are then computed and differential quality factor extracted. In this paper, for the first time using recently available vector network analyzer capable of delivering true differential signal, PNA-X from Agilent, we will verify this hypothesis discussing dedicated differential test structure and methodology. Moreover, from the modelling side new perspective will be proposed in order to take advantage of mixed mode S-parameters to extract new parameter such as self mutual inductance.

## II: Test Structures Definition

As previously said, up to now symmetrical centre tapped inductors have been characterized using conventional 2 ports test-structure suitable for probing with GSG RF-probes [6] as shown in Fig. 1. The four ground pads are connected together and to the patterned ground shield beneath the entire inductor structure. “Open” and “short” dummy structures [7] are used to de-embed the bondpad and interconnect parasitics.

Moving to true differential characterization, we have to modify the test structure in order to perform 2 port measurements using ground-signal-ground-signal-ground (GSGSG) RF probes. The dedicated test structure we propose for this purpose is presented Fig. 2. One major advantage of this new test structure in comparison with the conventional one is compacity. Using GSGSG RF probes enable to reduce the area consumed by pads and also to reduce interconnect which is a key advantage from de-embedding point of view. Moreover, the centre tapped is also connected to ground as it would be the case in a differential circuit.

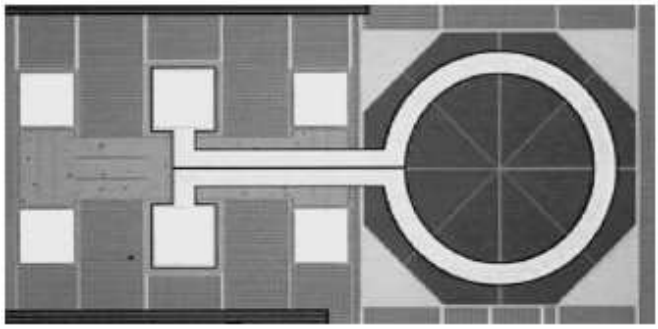


**Fig. 1**: Chip photograph of the on-wafer inductor test-structure in two-port ground-signal-ground configuration [5]

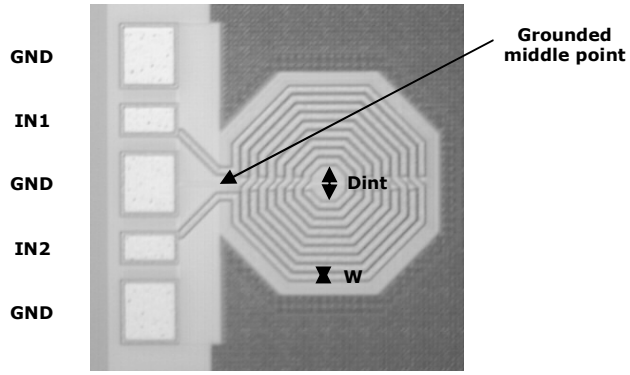


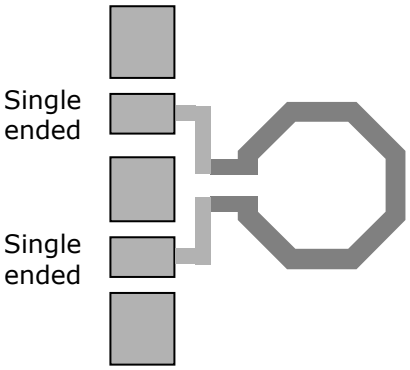

**Fig. 2**: Chip photograph of the proposed on-wafer inductor test-structure in two-port GSGSG configuration

## III: BALANCED CHARACTERIZATION

### *A. Pure Balanced and Build-in mixed Mode*

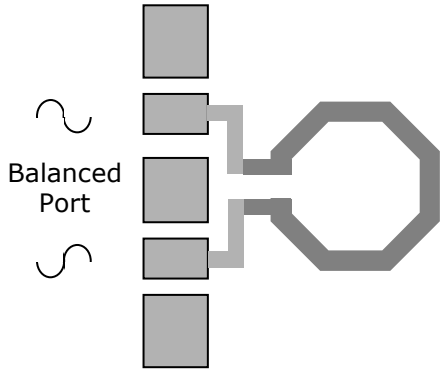
We perform the characterization of differential inductor with 4 ports PNA-X network analyzer from Agilent and Cascade Microtech Differential RF probe. This pure mode system measures the response of the circuit in the fundamental modes of operation: Balanced and Common mode. Two types of characterization have been performed: the single ended response (Figure 3) was measured and then the pure balanced mode response (Figure 4). In the first case, we convert single scattering parameters to build-in mixed mode scattering parameters. In the second case, we obtain directly the mixed S-parameters.

978-1-4244-4259-1/09/$25.00 ©2009 IEEE




$$[S]_{single} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}$$

**Fig. 3**: Schematic for two ports Single Ended measurement




$$[S]_{mixed} = \begin{bmatrix} S_{dd} & S_{dc} \\ S_{cd} & S_{cc} \end{bmatrix}$$

**Fig. 4**: Schematic for Mixed Mode measurement

*B. Method of de-embedding*

De-embedding of Pure Mode scattering parameters is not today clearly defined: there is no generic method without any assumption of linear behaviour [8]. For this work, we make hypothesis that interconnections are passives. Then parasitics only add current and voltage over the Pure modes. The de-embedding could be treated as follow with SHORT and OPEN structures.

| Pure mode S-parameters: | Build in mixed mode S-parameters: |
|---|---|
| 1) Open correction of Short | 1) Open correction of Short |
| $Y^*_{short\_cor} = Y^*_{short} - Y^*_{open}$ | $Y_{short\_cor} = Y_{short} - Y_{open}$ |
| 2) Open correction of the device | 2) Open correction of the device |
| $Y^*_{ind\_cor} = Y^*_{ind} - Y^*_{open}$ | $Y_{ind\_cor} = Y_{ind} - Y_{open}$ |
| 3) Short correction of the device | 3) Short correction of the device |
| $Y^*_{ind\_deemb} = ztoy\,(Z^*_{ind\_cor} - Z^*_{short\_cor})$ | $Y_{ind\_deemb} = ztoy\,(Z_{ind\_cor} - Z_{short\_cor})$ |
| | 4) Computation of build in mixed mode S-parameters |

One must take care that de-embedding with Z and Y mixed matrix (Matrix with (*) indicates mixed matrix) could not be easily linked to an equivalent circuit since electrical and physical ports are no more the same.

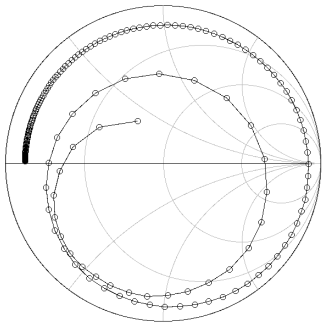
After de-embedding procedure, we obtain the S mixed mode parameters corrected from the interconnect lines and pad accesses. For a reciprocal and symmetrical device, the matrix simplifies to:

$$[S]_{mixed} = \begin{bmatrix} 0.5x(S_{11} - S_{12} - S_{21} + S_{22}) & 0 \\ 0 & 0.5x(S_{11} + S_{12} + S_{21} + S_{22}) \end{bmatrix}$$

The fact that the $S_{dc}$ and $S_{cd}$ are equal to zero shows no conversion from common mode to differential and inversely. With the same hypothesis concerning our device we can extract the mixed-mode Y-matrix and Z-matrix.

*C. Comparison between Pure Balanced and Build-in mixed Mode*

The following figure highlight the comparison of Pure Balanced mode and Build in-mixed mode measurement with PNA-X, up to 26 GHz, using the de-embedding procedure presented in the previous section.



○ $S_{11}$ Pure Balanced Mode
- $S_{11}$ Build-in mixed Mode

Freq. [0.1 GHz – 26 GHz]

**Fig. 5**: Pure Balanced Mode (○) versus Build-in mixed Mode (-) $S_{11}$ versus frequency (nbt = 9, Dint = 53 µm, w = 5 µm)

The results of pure balanced mode and build-in mixed measurements show a good agreement between two techniques of characterization. From this study, we conclude that we are going to use build in mixed mode, rather than pure balanced mode given that with the first mode we obtain only one equivalent scheme for the differential inductor and measurements are faster. Time consuming is preponderant in aiming scalable model extraction.

Even if those results were expected, it confirms for the first time all the hypothesis made up to now concerning the characterization and modelling of symmetrical centre tapped inductor. Moreover, as we will discuss in the following section, pure balanced mode capability open the way for new characterization and modelling validation strategy.

## IV : Modeling Strategy using Mixed-Mode S-Parameter to improve Model Parameter Extraction

Up to now, build in mixed mode S-parameters have been used only to compute the differential impedance of the inductor and then to extract the differential quality factor. From modelling point of view, having access to both common mode and differential S-Parameters could help to

extract new model parameter. The main one should be the self mutual inductance. When using a differential inductor in a circuit simulation, we have to introduce a mutual inductance the compact model in order to take into account any unbalanced.

In mostly used differential inductor compact model, people do not extract any mutual inductance experimentally, since there is no way to extract this parameter using single ended measurements. We can try to access to this parameter using electromagnetic simulation but generally the accuracy of the simulation is in the order of the accuracy of silicon based model delivered by foundry. Then, up to now it has been impossible to check experimentally mutual inductance value.

In order to overcome this limitation, we propose here an innovative way to extract the self mutual inductance of any kind of inductor using build in mixed mode S-parameters computed from 2 ports single ended measurements.

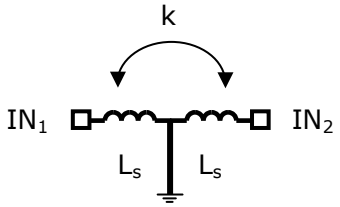



**Fig. 6**: Simple symmetrical centre tapped inductor model with a self mutual inductance

Using the simple inductor model presented Fig. 6, when applying a balance signal we have:

$$L_{dd} = \frac{\mathrm{Im}(stoz(S_{dd},100\Omega))}{2\times\Pi\times freq.} = 2\times L_s\times(1+k)$$

This parameter could be extracted using $S_{dd}$ coming from the mixed mode S-parameters. When applying a common mode signal we have:

$$L_{cc} = \frac{\mathrm{Im}(stoz(S_{cc},25\Omega))}{2\times\Pi\times freq.} = 2\times L_s\times(1-k)$$

This parameter could this time be extracted using $S_{cc}$ coming from the mixed mode S-parameters. Combining those two equations we have:

$$k = \frac{L_{dd}-L_{cc}}{L_{dd}+L_{cc}}$$

Then, we have now an easy way to extract the self mutual inductance of symmetrical centre tapped inductor. One example of extraction is presented Fig. 7.

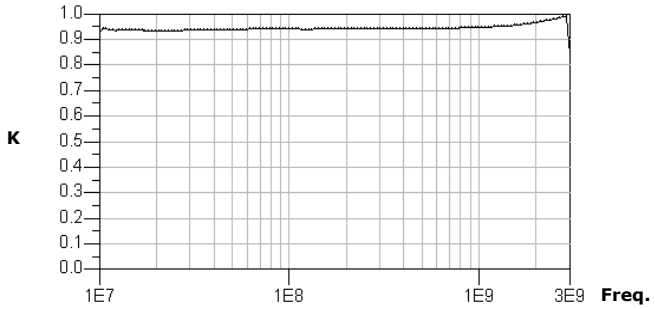



**Fig. 7**: Extraction results of parameter k using build in mixed mode S-parameters (nbt=9, w=5 µm, dint=80 µm)

Moreover, using the capability of PNA-X to ensure any phase difference (and not only the 180° related to differential mode) we are now able to verify that the new compact model taking into account the self mutual inductance is able to capture any unbalances.

## V: Conclusion

This paper presents a comparison between a pure mode and build in balanced characterization of inductor. As predicted by linear electronic theory, the true mode matrix is equal to the build in matrix for small signal analyses. Moreover, we have proposed innovative test structure, de-embedding techniques and parameter extraction methodology which take advantage of mixed mode S-parameters in order to determine experimentally the self inductance mutual. From modelling point of view, it opens the door for new compact model in order to be able to check and deal with any unbalance in the simulation.

## Acknowledgement

The authors would like to thank Emmanuel Rosello for helpful technical support.

## References

[1] G. Dambrine et al., "High frequency low noise potentialities of down to 65nm technology nodes MOSFETs", GaAs Symposium, 2005, Paris.

[2] J. N. Burghartz et al., "RF potential of a 0.18- m CMOS logic device technology," IEEE Trans. Electron Devices, vol. 47, pp. 864–870, Apr. 2000.

[3] C. H. Doan et al., "60 GHz CMOS radio for Gb/s wireless LAN," in IEEE RFIC Symp. Dig., June 2004, pp. 225-228.

[4] M. Danesh et al., "A Q-factor enhancement technique for MMIC inductors," in Proc. IEEE MTTS and RFIC Symp., pp. 217–220.

[5] L. TIiemeijer et al., "Record Q Symmetrical Inductors for 10 GHz LC VCOs in 0.18 um Gate Length CMOS", IEEE EDL, vol. 23, n° 12, December 2002.

[6] R. J. Havens et al., "Impact of probe configuration and calibration techniques on quality factor determination of GHz level on-wafer inductors," in Proc. ICMTS, 2002, pp. 19–24..

[7] M. A. C. M. Koolen, J. A. M. Geelen, and M. P. J. G. Versleijen, "An improved de-embedding technique for on-wafer high-frequency characterization," in Proc. BCTM, 1991, pp. 188–191.

[8] W.R. Eisenstadt, IEEE Microwave and Guided Wave letters, Vol.9, No.1, January 1999, pp. 25-27.

# In-Situ Silicon Integrated Tuner for Automated On-Wafer MMW Noise Parameters Extraction using Multi-Impedance Method for Transistor Characterization

**Y. Tagro[1,2], D. Gloria[1], S. Boret[1], Y. Morandini[1], G. Dambrine[2]**

[1]STMicroelectronics, Technology R&D – TPS, Crolles F – 38926, France
Tél : 33 34 76 92 66 84 ; E-mail : yoann.tagro@st.com

[2]Institut d'Electronique, de Microélectronique et de Nanotechnologie (IEMN),
Villeneuve d'Ascq Cedex 59652, France

***Abstract* — In this paper, for the first time, Silicon integrated tuner is presented aiming silicon transistor (HBT, MOSFET) MilliMeter Wave (MMW) noise parameters (NFmin, Rn, Γopt) extraction through multi-impedance method. This Tuner is directly integrated in On-wafer tested transistor test structure. Design, electrical simulation and MMW measurement of the Tuner are described showing capability from 60GHz up to 110GHz for CMOS and BiCMOS sub 65nm technologies characterization. |Γ| of 0.88 have been achieved at the DUT input in the considered frequency range and Tuner insertion losses are less than 20 dB.**

***Index Terms* — Active devices, transistors, HBT, MOSFET, in-situ lab., Impedance tuner, noise microwave measurement, multi-impedance, varactor, transmission lines, cold FET.**

## I. Introduction

In the last years, High Frequency (HF) performances of Si devices (CMOS: MOSFETs and BiCMOS: HBTs) have considerably increased and recent Silicon technologies provide dynamic performances (cut-off frequencies higher than 230GHz and $NF_{min}$ lower than 2dB@60GHz) very attractive for MMW-band mass market applications (60GHz WLAN & WPAN, Wireless HDMI, Wireless USB, 77GHz automotive radars, …). One of the key issues to optimize the technology and to develop accurate electrical model is the HF characterization and above all the transistor HF noise parameters extraction. It is well known in the theory of linear noisy networks that a complete characterization of the noise in a linear two-port at one frequency requires the knowledge of the four noise parameters $NF_{min}$, $R_n$, Real($\Gamma_{opt}$), Imag($\Gamma_{opt}$) [1] where $\Gamma_{opt}$ is the optimum reflection coefficient provided by the source admittance $Y_{opt}=G_{opt}+jB_{opt}$. Therefore, the challenge is to characterize these 4 noise parameters of the device up to 110GHz on DUT having $\Gamma_{opt}$ very different to 50Ω and $NF_{min}$ lower than 2dB in that frequency range. Two extraction methods are well known for the 4 noise parameters measurement: $NF_{50}$ method in case of MOSFET [3] or multi-impedance method for Bipolar and MOSFET [2]. The 2nd method is based on the use of an impedance synthesizer (Tuner) that is Off-wafer and external to the On-wafer test fixture. This paper presents the development of an in-situ tuner integrated on the On-wafer test structure and used to apply multi-impedance method.

Firstly, existing classical noise extraction methods and their limits are described. Secondly, we describe the proposed approach and its strengths for Silicon device On-wafer automated measurement. Finally, we present tuner results from simulation to measurement.

## II. $NF_{50}$ and Mutli-Impedance Noise Extraction Method And Their Limitations

### *A. $NF_{50}$ Method:*

Only available for FET devices it is an analytical method assisted by a physical-based noise model [3]. The aim of this technique is to not use systematically a multi-impedance test bench. This kind of approach allows to use only noise measure on 50Ω impedance and the scattering S parameter of the DUT. The scattering S parameter is used to extract the element of the DUT equivalent model, the noise factor on 50Ω allows to obtain the noise sources of the device. This method is based on (1) and assumes that $R_n$ is frequency independent and that $|Y_{opt}|^2$ varies as $\omega^2$.

$$NF_{50} = 1 + R_n.G_0 + \frac{R_n}{G_0}.\left(2.G_0.G_{cor}+\left|Y_{opt}\right|^2\right) \quad (1)$$

These 2 assumptions are verified for MOSFET but not for Bipolar and that is why the proposed lab in-situ method is based on the flow described on part B.

### *B. Multi-impedance method:*

Introduced by R.Q Lane at the end of the 60th year [2], the multi-impedance technique consists in the extraction of the 4 noise parameters by the measurement of 4 noise factors for 4 different source admittances $Y_S=G_S+jB_S$.

The method is based on Fig.1 setup using (P1, P2) reference plane and (2) transformed into (3) which is linear regarding 4 new parameters (A, B, C and D) that can be obtained by the measurement of a minimum set of 4 noise factors for 4 different admittances ($Y_S$).

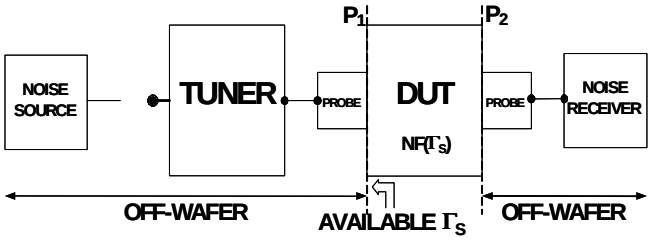



Fig. 1. Multi-impedance method test bench [Tuner Off-WAFER]

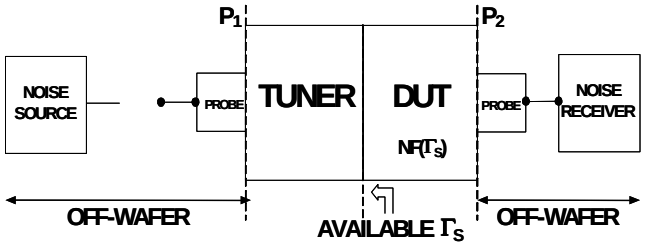



Fig. 2. Multi-impedance method test bench [Tuner On-WAFER]

Thanks to (A, B, C, D) determination, and using (4), we can determine the 4 noise parameters. Nevertheless, for extraction accuracy reason, we approximately use "N=80" different source admittances ($Y_S$), and deduce A, B, C and D by minimizing the error $\varepsilon$ defined by (5):

$$NF = NF_{\min} + \frac{R_n}{G_s}\left|Y_s - Y_{opt}\right|^2 \tag{2}$$

$$NF = \left[A + B\left(G_S + \frac{B_S^2}{G_S}\right) + \frac{C}{G_S} + \frac{DB_S}{G_S}\right] \tag{3}$$

$$4 \quad Noise \quad Parameters = \begin{cases} NF_{\min} = A + \sqrt{4BC - D^2} \\ R_n = \dfrac{NF_{\min} - A}{2.G_{opt}} = B \\ G_{opt} = \dfrac{\sqrt{4BC - D^2}}{2.B} \\ B_{opt} = \dfrac{-D}{2.B} \end{cases} \tag{4}$$

$$\varepsilon = \frac{1}{2}\sum_{i=1}^{N} \frac{1}{NF_i^2}\left[A + B\left(G_{Si} + \frac{B_{Si}^2}{G_{Si}}\right) + \frac{C}{G_{Si}} + \frac{DB_{Si}}{G_{Si}} - NF_i\right]^2 \tag{5}$$

The main limitation of this method is the range of |Γ| achievable in the DUT reference plane due to losses between external tuner and DUT input (cables, connectors, RF probes) mainly in MMW frequency range. The proposed solution is based on multi-impedance concept but uses integrated tuner in the test structure On-wafer.

## III. In-Situ Tuner Design

The measurement concept is illustrated on Fig.2 using (P1, P2) reference plane.

### *A. Tuner Concept*

The architecture of the tuner uses a cold-nMOS (as variable resistance) in series with a varactor (as variable capacitance) and a TL. Fig.3 shows an electrical schematic of this impedance tuner. All used components are passives.

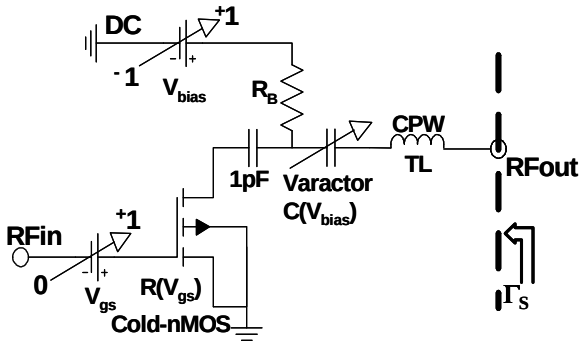



Fig. 3. Electrical schematic of the in-situ tuner in 65nm HR SOI technology.

First of all, to know impedance to be synthesized, a large set of MOS transistors in the tested technology has been studied through electrical model in term of stability circles, optimal source reflection coefficient for $NF_{min}$, and available gain in the concerned frequency range. (see Fig.9).

### *B. Coplanar TL in the 65nm HR SOI technology.*

The first prototype has been achieve on STMicroelectronics 65nm HR SOI technology on High Resistivity (HR) substrate providing 6 metallization levels for analog applications and more for digitals [4] as shown on Fig.4.

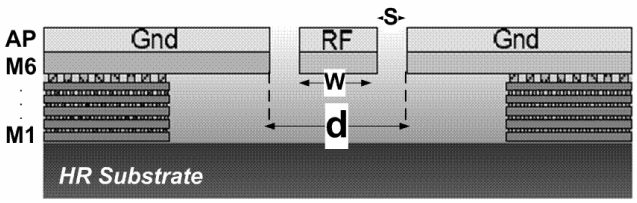



Fig. 4. representation of the Coplanar Transmission Line in 65nm HR SOI BEOL [5], and overview of the 6 metallization level

The used TL is a coplanar one described in Fig.4. The characteristic impedance ($Z_C$) of this coplanar line is near 50Ω and has been obtained for the following dimension

(d=70µm and W=26µm [6]). Fig.5 shows its frequency response.

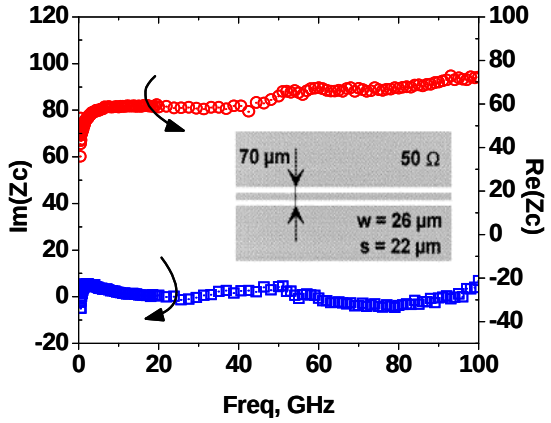



Fig. 5. Experimental frequency response on 50-Ω CPW TL

*C. Varactor used as variable capacitance.*

The next passive component used to design the tuner is a varactor based on $N^+$poly/Nwell structure. The device is made of 5 poly fingers in parallel. Each finger has 0.35µm length and 3µm width.

Fig.6 shows the frequency and the bias response of this varactor.

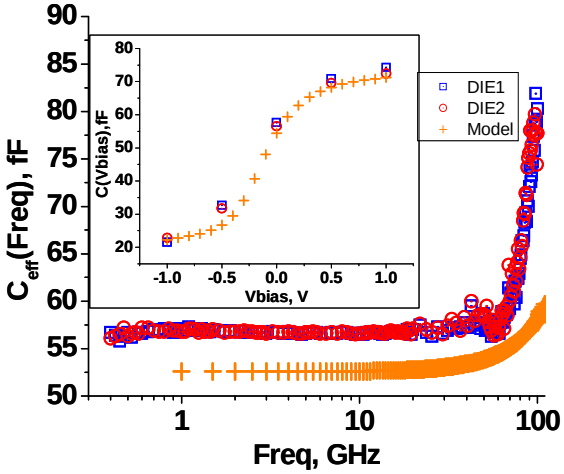



Fig. 6. C(Vbias) versus Vbias @ 60GHz and C(Freq) versus Log(Freq) @0V.

The criterions to optimize the choice of this component are:

(i). the cut-off frequency must be higher than the maximum frequency range (in our case: 100GHz).

(ii). The tuning range must be high (~3.5 in our case).

*D.. Cold-FET used as variable resistance.*

Cold-FET is obtained using MOS transistor with floating drain bias (Fig.7). It's made of 80 poly fingers in parallel contacted on both side and each finger has a length of 0.12µm and a width of 0.5µm.

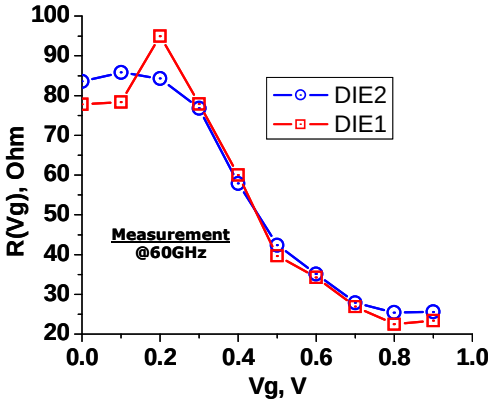



Fig. 7. Measured, Cold-FET versus bias.

## IV. Experimental Results

In this section, we present the experimental data of this study. The measurements are done in 400MHz to 100GHz frequency range even if the used band is 60GHz to 100GHz. These measurements are done on two die of 300mm wafer in 65nm HR SOI technology.

*A. In-situ Tuner measurement results.*

The first Tuner has been achieved with the use of all previous lumped components described above. We can identify these components on the Tuner's layout (Fig.3, 8).

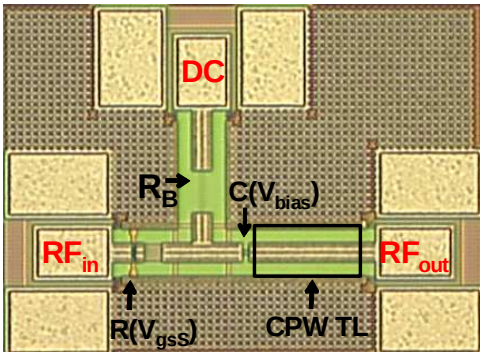



Fig. 8. In-Situ Tuner layout.

On the Fig.9 we present the results of synthesized admittance $Y_S$ with the achieved tuner. 3 zones can be identified corresponding to 3 different TL lengths. The lengths have been chosen to move on the smith chart avoiding instability area, and moving across different noise figure circle for the nMOS of the studied technology. Inside each zone obtained for a given TL length, the different admittance are obtained thanks to $R(V_g)$ (Cold-FET) and $C(V_{bias})$ (Varactor) bias controlled.

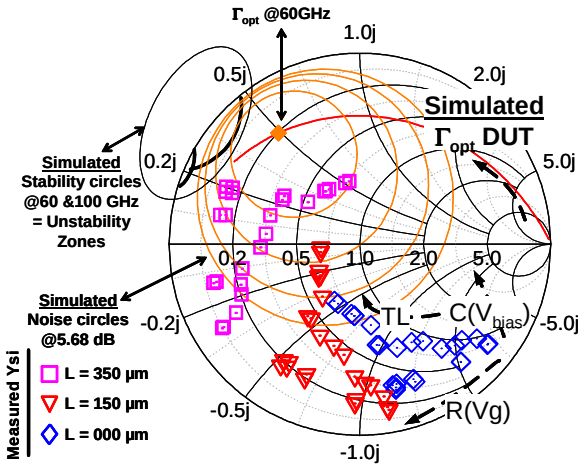



Fig. 9. Impedance synthesis with In-Situ Tuner and 65nm HR SOI nMOS (40*0.06um^2) simulated noise characteristics $V_{gs}$=0.72V&$V_{ds}$=1.2V.

On Fig. 9, based on tuner specifications study described on III.A, are plotted the stability circle for 60 and 100GHz, noise circle @60GHz with 1dB step, and the $\Gamma_{opt}$ from 400MHz up to 100GHz of a typical 65nm HR SOI nMOS (0.06*40um^2). These added characteristics show that the available $Y_S$ achieved with Tuner provide a good potential for accurate noise extraction on 65nm devices.

In addition of (Fig.9), Fig.10 and Fig.11 shows the measurement impedance coverage up to 100GHz, showing good coverage for each frequency.

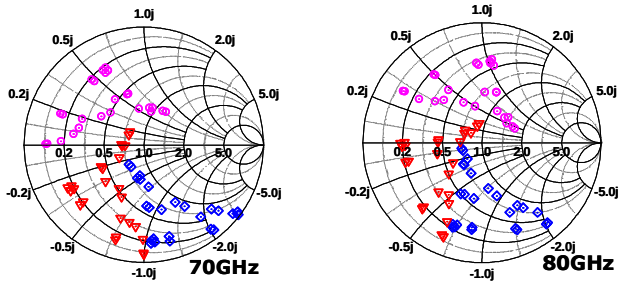



Fig. 10. Impedance area coverage (70 & 80GHz), same legend as Fig.9.

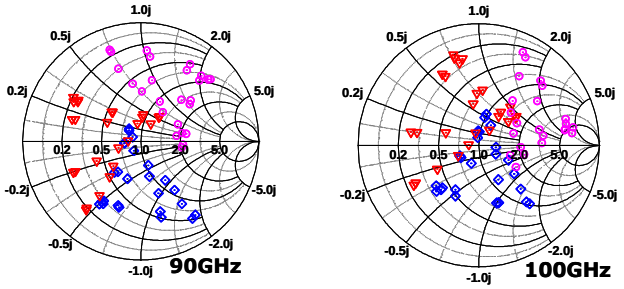



Fig. 11. Impedance area coverage (90 & 100GHz), same legend as Fig.9.

A nMOS with (40*0.06um^2) geometry has been measured with this tuner at Vgs=0.72V and Vds=1.2V.Obtained result at 78GHz are NFmin=2dB, Rn=25Ω and Γopt~0.85/110°.

*B. Perspectives of Tuner improvement*

As shown in Fig.12, we can see a good agreement between Tuner measurements and electrical simulations. Nevertheless, we observe insertion losses around 20dB in the 60-100GHz range. Reducing the Cold-FET resistance would improve losses as shown in the same figure. This losses reduction will allow higher $\Gamma(Y_S)$ at the DUT input. In addition reducing $C_{gs}$ of the cold-FET will also improve losses.

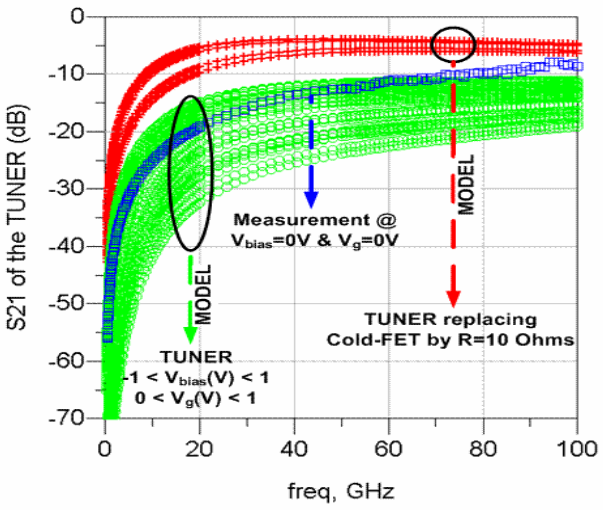



Fig. 12. Tuner $S_{21}$ simulation versus measurements “Effect of Cold-FET resistance value”.

Fig.13 shows the methodology flow to fix the Tuner specification. We use the measured admittances shown in Fig.9-11 in in-house developed ADS® multi-impedance tool

dedicated to this development and applying them directly at the input of Si electrical model device (CMOS: MOSFETs and BiCMOS: HBTs) to obtain their 4 noise features.

The noise features obtained using achieved $Y_S$ on this tool at 78GHz are:

$NF_{min}$=2.05 dB, $R_n$=18Ω and $\Gamma_{opt}$=0.70/121°

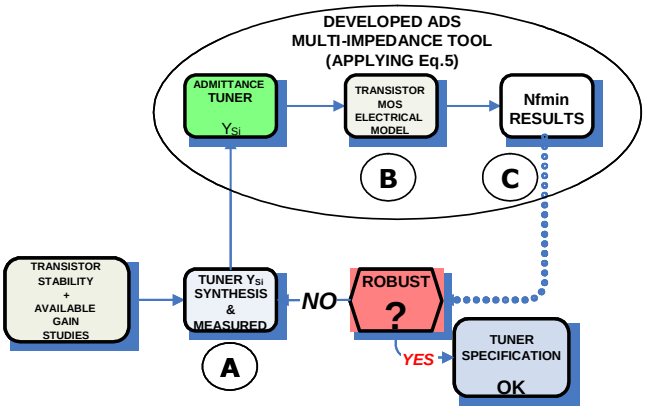



Fig. 13. Methodology flow to fix TUNER specifications using ADS in-house multi-impedance tool

We can improve this tool box adding tuner losses in (A) block to take the losses as parameters to be optimized.

For that, maximum noise figure of 12dB of DEVICE+TUNER will be considered between (B) and (C) to define maximum allowed losses in the tuner according to tested transistor gain.

## V CONCLUSION

A new way to extract the high frequency noise parameters $NF_{min}$, $R_n$, $G_{opt}$ and $B_{opt}$ into the frequency range from 60GHz to 110GHz is presented on this paper. This method is based on the multi impedance technique using an in-situ tuner directly integrated on the test structure of HBT or MOSFET devices in sub 65nm technologies. The design of the tuner is based on Cold-FET, and varactor variable with biasing and associated to TL. Obtained electrical results on synthesized impedance are shown and compared to electrical simulations demonstrating a good agreement. A nMOS with (40*0.06um^2) geometry has been measured with this tuner at Vgs=0.72V and Vds=1.2V. Obtained result at 78GHz are NFmin=2dB, Rn=25Ω and Γopt~0.85/110°.

## ACKNOWLEDGEMENT

This work has been carried out within the STMicroelectronics Center R&D, and the authors wish to thank the staff of the 300mm Si facilities in ST Crolles for device fabrication

## REFERENCES

[1] H. Rothe. W. Dahlke, "Theory of noise fourpoles," *2003 IEEE Proc. Of the IRE* , pp. 811-818, June 1956.

[2] R. Q. Lane, "The determination of device noise parameters," *Proc. Of the IEEE*, vol. 57, pp. 1461-1462, August 1969.

[3] G. Dambrine and all, "A new method for On-Wafer Noise Measurement," *IEEE Trans. On Microwave Theory & Tech.*, vol. 41, no. 3, pp. 375-381, March 1993.

[4] F. Gianesello, "1.8 dB insertion loss 200GHz CPW band pass filter integrated in HR SOI CMOS Technology," *IEEE Trans. On Microwave Theory & Tech.*, pp. 453-456, 3-8 June 2007.

[5] S. Montusclat, "Design for Millimeter-wave applications in Silicon Technologies," *IEEE 33rd ESSCIRC*, pp.464-471, 11-13 September2007.

[6] S. Boret, "Wide- and narrow-band bandpass coplanar filters in the W-frequency band," *IEEE Trans. On Microwave Theory & Tech.*, vol. 51, no. 3, pp. 784-791, March 2003.

# Four point probe structures with buried electrodes for the electrical characterization of ultrathin conducting films

A.W. Groenland[1], R.A.M. Wolters[1,2], A.Y. Kovalgin[1] & J. Schmitz[1]

[1] MESA+ Institute for Nanotechnology, Chair of Semiconductor Components
University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands.
Tel +31 (0)53 489 2645, Fax +31 (0)53 489 1034
Email: A.W.Groenland@utwente.nl

[2] NXP-TSMC Research Center, High Tech Campus 4, 5656 AA Eindhoven, The Netherlands

## ABSTRACT

Test structures for the electrical characterization of ultrathin conductive (ALD) films are presented based on buried electrodes on which the ultrathin film is deposited. This work includes test structure design and fabrication, and the electrical characterization of ALD TiN films down to 4 nm. It is shown that these structures can be used successfully to characterize sub 10 nm films.

## INTRODUCTION

Conducting thin films (sub 10 nm) have interesting applications in the field of diffusion barriers for copper metallization and phase change memories [1-4]. Electrical properties can be measured using the four probe method [5, 6]. For these thin films, van der Pauw structures (vdPs) and Greek Crosses (GCs) are commonly used [7]. However, making electrical contacts to these thin films, using planar technology, is extremely difficult; e.g. etching a via for the contact to such a film requires a very high selectivity. In practice, this is not possible. Enderling and co-workers proposed to use suspended Greek Crosses to overcome this problem. However, this is only suitable for films with a poor step coverage (e.g. deposited via physical vapour deposition (PVD)) [8].

In this work special test structures are presented to measure the electrical properties of thin films in a controlled way. Electrodes are fabricated, which are buried in a planarized dielectric film. A thin film is deposited on top of the electrodes by atomic layer deposition (ALD). The thin film is characterized using the predefined electrode structures, such as the aforementioned vdPs and GCs. The work presented here includes test structure design and fabrication. It is demonstrated that these structures can be used successfully to characterize 7 nm and 4 nm thick ALD TiN films.

## TEST STRUCTURE FABRICATION

Square structures are etched in a layer of 0.5 µm thermally grown $SiO_2$ on top of a standard silicon wafer using wet chemical etching (Fig. 1a). After an additional oxidation step to insulate the silicon substrate, a layer of 70 nm TiW is sputtered and patterned to make the electrodes and connections (Fig. 1b). Subsequently a layer of 1 µm PECVD $SiO_2$ is deposited (Fig. 1b) and the structure is planarized using chemical mechanical polishing (CMP) (Fig. 1c). On this surface the thin ALD TiN film is deposited which is passivated by in situ ALD $Al_2O_3$ and ex situ PECVD $SiO_2$ layers (Fig. 1d). After patterning the ALD TiN and passivation layers (Fig. 1e), vias are etched towards the buried electrode connections and filled with aluminium to become realize contact pads (Fig. 1f). In this process, structures such as vdPs, GCs, collinear probes and contact strings have been realized. The schematic design and exact device dimensions of a vdP and GC are shown in Fig. 2. An example of a realized van der Pauw structure is shown in Fig. 3.

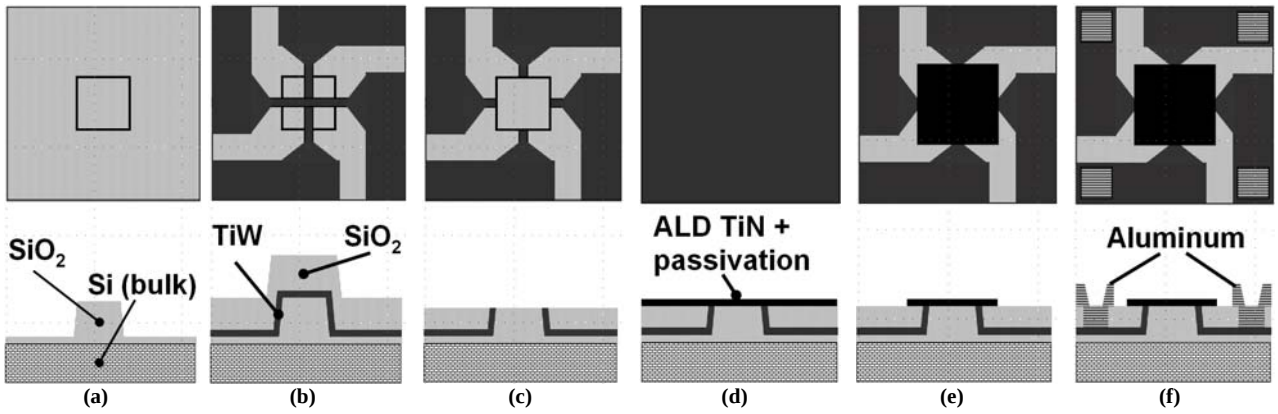
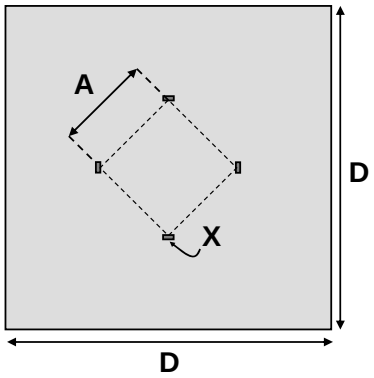


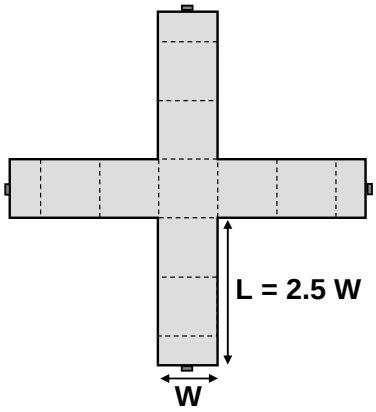

Fig. 1. Fabrication scheme for a van der Pauw structure (top view (top) and cross-section (bottom)).

| A [µm] | D [µm] |
|---|---|
| 10 | 172 |
| 25 | 357 |
| 50 | 357 |
| 75 | 500 |
| 100 | 500 |
| 300 | 530 |



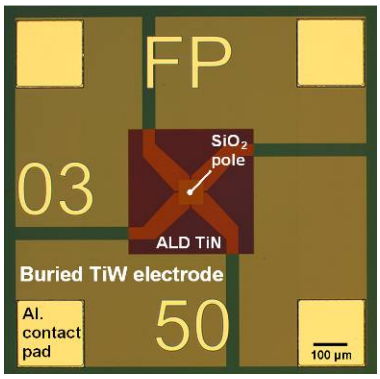

(a)

| W [µm] |
|---|
| 10 |
| 25 |
| 50 |
| 75 |
| 100 |




(b)

Fig. 2. Schematic design of a van der Pauw (a) and Greek Cross (b) device. The four contacts of the van der Pauw are situated at the corners of an imaginary square at a distance $A$ that is centred with respect to the ALD film. The ALD film is patterned as a square with dimension $D$. $X$ indicates the area of a single contact (2.5 µm × 0.1 µm for all devices). The Greek Cross is designed with the arm length $L$ equal to 2.5 times the width $W$. This reduces the error in the extracted sheet resistance to < 1% while maintaining maximum sensitivity and minimal Joule heating in the arms of the cross [9-11]. Dashed lines indicate the (imaginary) squares.



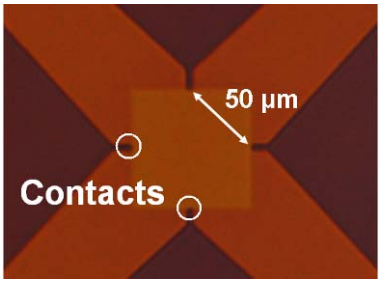
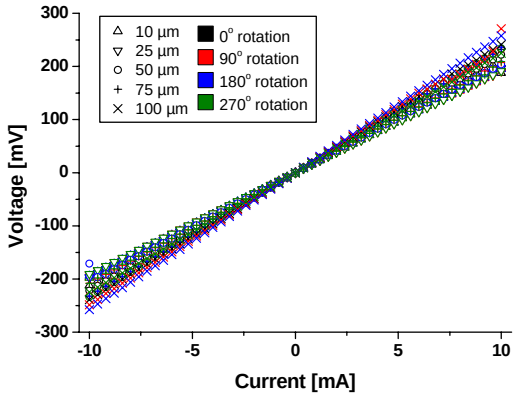
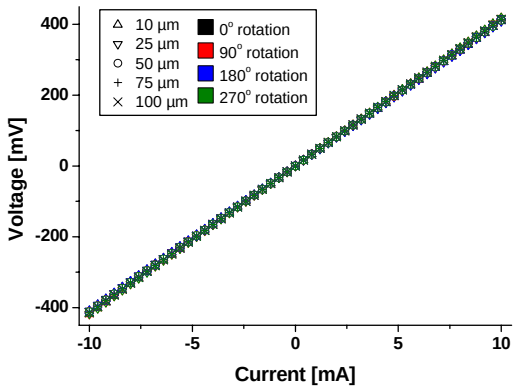




Fig. 3. Optical micrograph of a van der Pauw structure with a probe-to-probe distance of 50 µm. Whole structure (top) and close up of the electrodes (bottom).

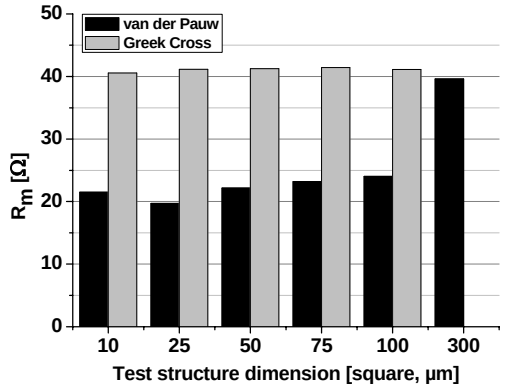



(a)

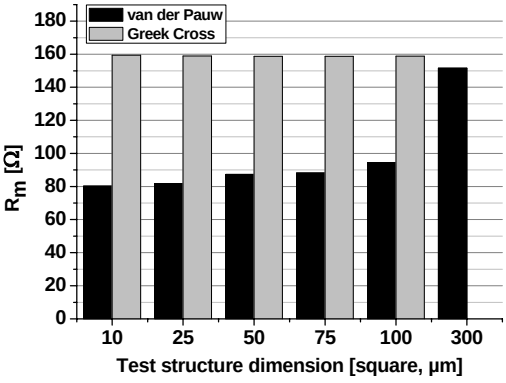



(b)

Fig. 4. Measured $V_m$ versus $I_m$ curves for van der Pauw (a) and Greek Cross (b) test structures for 7 nm ALD TiN. Dimensions refer to the probe-to-probe distance (van der Pauw) and the size of the central square of the Greek Cross.




(a)




(b)

Fig. 5. Extracted resistances ($V_m/I_m$) from van der Pauw and Greek Cross structures for 7 nm (a) and 4 nm (b) ALD TiN layers. Resistances are an average of the 4 measured orientations (see Fig. 4).

## EXPERIMENTAL

The layer thickness of the ALD TiN films is determined using a Woollam M2000 Spectroscopic Ellipsometer (SE) in the energy range 0.7-5 eV. Measurements were taken *in situ* directly after deposition. From the recorded SE data, the ALD TiN layer thickness is derived using a model containing the optical constants of all sub-layers. The ALD TiN layer thickness and the contact area of the electrodes are verified by HRSEM on cross-sections of the sample.

For the electrical characterization, *IV*-measurements were carried out using a HP4156B or Keithley 4200 precision semiconductor parameter analyser in combination with a Cascade Microtech probe station. For the measurements at elevated temperatures, the temperature controlled chuck of the probe station was used.

## RESULTS

### *A. Measurements at room temperature*

For a 7 nm ALD TiN layer, *IV*-curves are measured from van der Pauw (Fig. 4a) and Greek Cross structures (Fig. 4b) having probe-to-probe distances (vdP) or central squares (GC) in the range of 10 × 10 to 100 × 100 μm$^2$. All devices are measured over all 4 orientations. From (solely) the slope of both sets of *IV*-curves, the ('measured') resistance ($R_m = V_m/I_m$) is calculated and averaged over the 4 orientations. The results are shown in Fig. 5a. In Fig. 5b $R_m$ values are shown for a 4 nm ALD TiN layer. Values of $R_m$ values extracted from a van der Pauw device with a probe-to-probe distance of 300 μm are included in the graphs

For van der Pauw structures, a relatively large spread in the *IV*-characteristics (Fig. 4a) is observed for different device dimensions and also between different orientations of a single device. For Greek Crosses (Fig. 4b) this is not the case.

From Fig. 5 it is observed that resistances extracted from van der Pauw structures (except for the 300 μm device) are significantly lower than those extracted from Greek Crosses. A slight increase in $R_m$ is observed for larger devices. The reduced measured resistance may be related to the finite contact area (~2,5 × 0.1 μm$^2$) of the electrodes (i.e. not point-like contacts) in the van der Pauw structures [5, 13, 14].

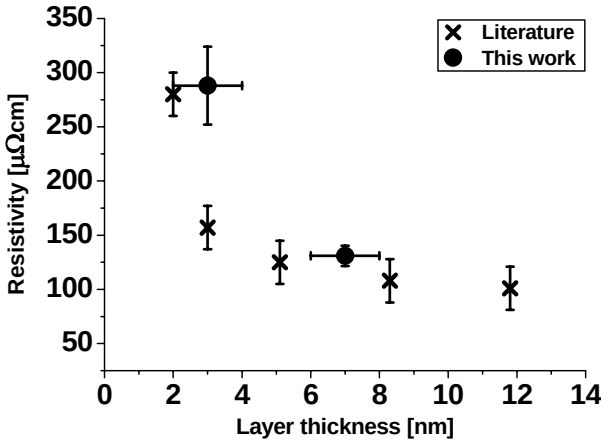



Fig. 6. Resistivity versus layer thickness for ALD TiN films. Literature values are reprinted from [12]. The uncertainty in our data is due to uncertainties in the layer thickness, as derived from spectroscopic ellipsometer (SE) measurements.

Furthermore, an increase is observed in $R_m$ extracted from van der Pauw structures for larger devices. This is most likely due to the fact that the edge of the ALD layer is situated further away from the electrodes for smaller devices, thereby violating one of van der Pauw's boundary conditions that the contacts should be at the circumference of the sample [5]. This is supported by measurements on the 300 × 300 μm$^2$ van der Pauw device in which the electrodes are close to the ALD layer edge: they yield virtually the same values for $R_m$ as extracted from Greek Crosses.

For Greek Cross devices, the sheet resistance ($R_\square$) can be calculated from $R_m$ using a correction factor of π/ln(2) [7]. The resistivity ($\rho$) is calculated from $R_\square$ and the layer thickness (averaged over all 5 devices). Values of 131 and 288 μΩcm are obtained for 7 and 4 nm ALD TiN respectively. These results are shown in Fig. 6 together with results from literature [12]. The error in $\rho$ originates from uncertainty in the ALD TiN layer thickness. Our extracted resistivity values are slightly higher than the literature values, but still realistic [2, 12, 15]. The difference is may be due to differences in material composition and/or the uncertainty in the layer thickness.

For van der Pauw devices it is more difficult to extract $R_\square$. As mentioned above, the contacts have finite size and are not placed at the circumference of the sample, so a correction factor larger than the standard correction factor of π/ln(2) has to be used. This factor cannot be derived easily in an analytical way, but might be obtained from finite element modelling [16]. This is beyond the scope of our present work.

### *B. Measurements at elevated temperatures*

Both van der Pauw and Greek Cross devices can be used for measurements at different temperatures to find the temperature coefficient of resistance (TCR) of the ALD TiN film. For a 7 nm ALD TiN film, resistance measurements in the range of 25 - 175 ºC are shown in Fig. 7. In Fig. 7a the measured resistance $R_m$ of a 75 μm van der Pauw and a 75 μm Greek Cross device are shown. A linear fit through the data can be described as:

$$R_m = aT + R_0 \qquad (1)$$

with $R_m$ the $V_m/I_m$ [Ω] and $T$ the temperature [ºC]. The temperature coefficient of resistance ($\beta$ or TCR) [ºC] can be extracted from the fit parameters $a$ and $R_0$ by rewriting (1):

$$R_m = R_0(1+\beta T) \rightarrow \frac{R_m}{R_0} = \beta T + 1 \qquad (2)$$

with $\beta = a/R_0$ [17].

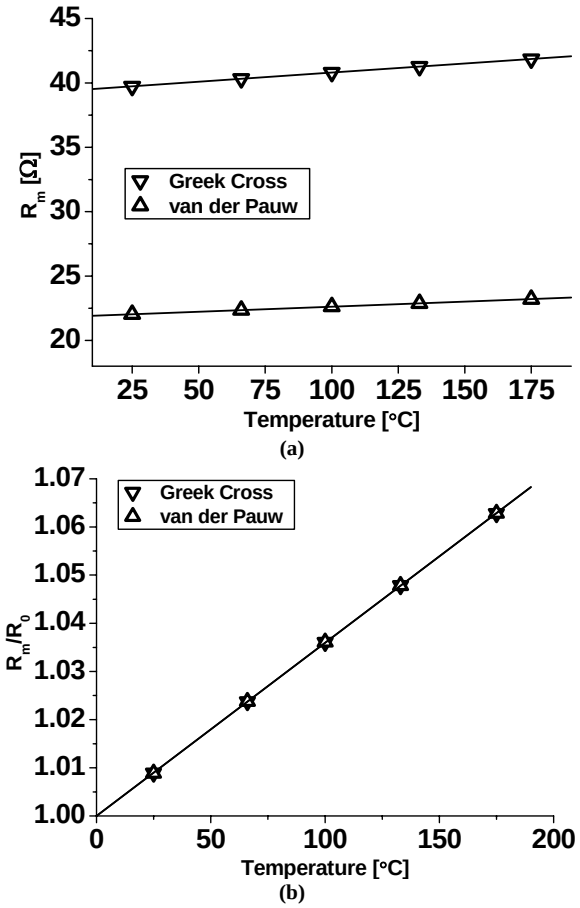



Fig. 7. Resistance versus temperature curves of 75 µm van der Pauw and 75 µm Greek Cross devices for a 7 nm ALD TiN layer. (a) the measured resistance ($R_m$) and (b) $R_m$ normalized on the (extrapolated) $R_m$ value at 0 ºC ($R_0$). Lines are linear fits through the data. The temperature coefficient of resistance (TCR) is extracted from the slope of (b) (see text).

In Fig. 7b $R_m/R_0$ is plotted versus temperature for the same devices as in Fig. 7a. Although the absolute $R_m$ values of the van der Pauw and the Greek Cross differ, it is observed that the normalized resistance values yield the same TCR ($3.59 \times 10^{-4}$ /ºC) for both devices.

In Fig. 8 TCR values for van der Pauw and Greek Cross devices with dimensions in the range of 10-300 µm are shown. It is observed that the spread in TCR values obtained from Greek Crosses is very low (within ~1% of their average value). For van der Pauw devices a larger spread (7% at maximum) in TCR values is observed. This spread might be related to a spread in device properties across the wafer; the 75 and 100 µm vdP & GC devices were positioned close to each other (neighbouring devices) and yield virtually equal TCR values (within 0.8% of their average value). All other devices were measured at different locations across the wafer (i.e. not within the same die). It appears van der Pauw devices are more sensitive to this spread in device properties than Greek Crosses.

The extracted TCR from Greek Cross devices of $3.5 \times 10^{-4}$ /ºC is roughly 40% lower than the literature value of $5.5 \times 10^{-4}$ /ºC, measured by Langereis and co-workers on an ALD TiN layer that is deposited under similar

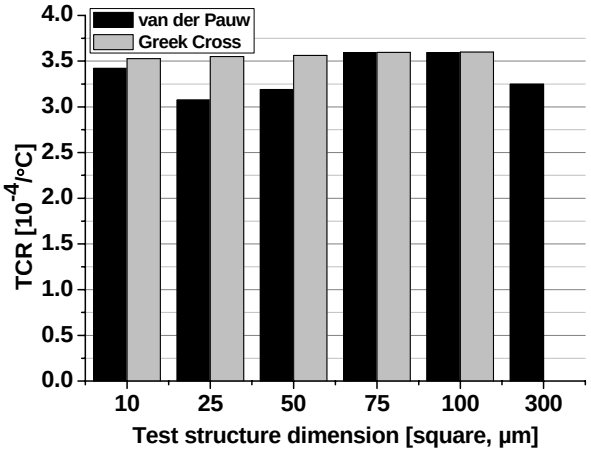



Fig. 8. Temperature coefficient of resistance (TCR) values of a 7 nm ALD TiN film for various van der Pauw and Greek Cross devices. Dimensions refer to the probe-to-probe distance (van der Pauw) and the size of the central square of the Greek Cross.

conditions [12, 18]. The difference is most likely due to a reduction of the TCR as a result of enhanced surface scattering which occurs for extremely thin films. Langereis and co-workers used a 33 nm thick film for the determination of their TCR which is significantly thicker than our 7 nm film [12, 18].

## CONCLUSIONS

Test structures for the electrical characterization of ultrathin conductive films are presented based on buried electrodes on which the ultrathin film is deposited. Electrical measurements on Greek Cross structures yield resistance values which are independent of the device dimensions ($10 \times 10$ to $100 \times 100$ µm$^2$) and the extracted values for the resistivity of 288 and 133 µΩcm for 4 and 7 nm ALD TiN layers, respectively, are realistic. Both van der Pauw and Greek Cross devices can be used for the extraction of the temperature coefficient of resistance (TCR) of ALD TiN. For a 7 nm ALD TiN layer, a TCR value of $3.5 \times 10^{-4}$ /ºC is found.

## ACKNOWLEDGMENT

The authors thank M. Hoekman (LioniX BV) for help with CMP processing, T. Aarnink (MESA+) for help with the ALD of TiN and the Dutch Technology Foundation STW for financial support of this project (07682).

## REFERENCES

[1] International Technology Roadmap for Semiconductors: 2007 Update, International Sematech, 2007. Available from: http://public.itrs.net.

[2] Ch. 2: Atomic layer deposition, in Handbook of thin film materials part 1: Deposition and processing of thin films (2002), vol. 1, edited by H. S. Nalwa, Academic Press.

[3] D. T. Castro, L. Goux, G. A. M. Hurkx, K. Attenborough, R. Delhougne, J. Lisoni, F. J. Jedema, M. A. A. t Zandt, R. A. M. Wolters, D. J. Gravesteijn, M. Verheijen, M. Kaiser, and R. G. R. Weemaes, "Evidence of the Thermo-Electric Thomson Effect and Influence on the Program Conditions and Cell Optimization in Phase-Change Memory Cells," in *2007 IEEE Electron Devices Meeting,* pp 315-318.

[4] J. Sarkar and B. Gleixner, Applied Physics Letters, Vol. 91, No. 23, 2007, pp. 233506-3.

[5] L. van der Pauw, Philips Research Reports, Vol. 13, No. 1, 1958, pp. 1-9.

[6] L. van der Pauw, Philips Technical Review, Vol. 20, No., 1958, pp. 220-224.

[7] M. G. Buehler and W. R. Thurber, Journal of The Electrochemical Society, Vol. 125, No. 4, 1978, pp. 645-650.

[8] S. Enderling, C. L. Brown, III, S. Smith, M. H. Dicks, J. T. M. Stevenson, M. Mitkova, M. N. Kozicki, and A. J. Walton, Semiconductor Manufacturing, IEEE Transactions on, Vol. 19, No. 1, 2006, pp. 2-9.

[9] S. Enderling, M. H. Dicks, S. Smith, J. T. M. Stevenson, and A. J. Walton, "Thermal design considerations for Greek cross test structures," in *2003 IEEE International Conference on Microelectronic Test Structures,* pp 8-13.

[10] S. Smith, "Sheet Resistance and Electrical Linewidth Test Structures for Semiconductor Process Characterisation" (2003), Ph.D. thesis, University of Edinburgh, Scotland.

[11] Y. Sun, J. Shi, and Q. Meng, Semiconductor Science and Technology, Vol. 11, No. 5, 1996, pp. 805-811.

[12] E. Langereis, S. B. S. Heil, M. C. M. van de Sanden, and W. M. M. Kessels, Journal of Applied Physics, Vol. 100, No. 2, 2006, pp. 023534.

[13] W. Versnel, Solid-State Electronics, Vol. 21, No. 10, 1978, pp. 1261-1268.

[14] M. Cornils, M. Doelle, and O. Paul, IEEE Transactions on Electron Devices, Vol. 54, No. 10, 2007, pp. 2756-2761.

[15] S. B. S. Heil, E. Langereis, F. Roozeboom, M. C. M. van de Sanden, and W. M. M. Kessels, Journal of the Electrochemical Society, Vol. 153, No. 11, 2006, pp. G956-G965.

[16] N. M. P. Guillaume, M. W. Cresswell, R. A. Allen, S. Everist, and L. W. Linholm, "Comparison of sheet-resistance measurements obtained by standard and small-area four-point probing," in *1999 IEEE International Conference on Microelectronic Test Structures,* pp 62-66.

[17] CRC Handbook of Chemistry and Physics (1987), ed, edited by CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida, 33431, USA.

[18] E. Langereis, private communication, 2009.

# Metal and Dielectric Thickness: a Comprehensive Methodology for Back-End Electrical Characterization

L. Bortesi[1], L. Vendrame[1]

[1]Numonyx, R&D - Technology Development
Via C. Olivetti 2, 20041, Agrate Brianza, Italy

## ABSTRACT

Back-End-Of-Line (BEOL) process variation is becoming more and more important since technology is scaling down and increases its complexity. On-chip capacitances and resistances are strongly dependent on the BEOL geometrical configuration so it is really important to have an accurate characterization of the metal and dielectric thickness. Interconnect parasitic modelling by means of LPE tool (Layout Parasitic Extraction) or semi-analytic approximation can't neglect the impact of metal (dielectric) thickness variations. The focus of this work is to provide an accurate, simple and suitable for parametric testing methodology to electrically measure metal (dielectric) thickness, mandatory for a useful characterization and control of a technology.

## INTRODUCTION

Interconnect parasitic elements s are nowadays a limiting factor for the overall performance of the entire chip: RC delay, cross-talk and dynamic cross-talk must be carefully taken into account during the circuit design. Furthermore also their variations due to process spread must be considered and correctly evaluated [1, 2]. The proposed approach gives a good estimation of the metal (dielectric) thickness allowing a precise map of its distribution across the wafer. The starting point of the developed methodology is the capacitance *C* of an ideal parallel plate:

$$C = \frac{\varepsilon_0 \cdot \varepsilon_r \cdot A}{d} \qquad (1)$$

where $\varepsilon_0$ is the permittivity of the free space, $\varepsilon_r$ the relative dielectric constant, *A* the parallel plate area and *d* the distance between the electrodes, e.g. the distance between different metal layers. The basic idea is to design an interconnect load based on two metal levels which has a capacitance well approximated by (1) and combining it with an accurate capacitance measurement system to obtain an indirect measure of *d*. The synergic work between the interconnect structure and the capacitance measurement system represents the core of the methodology. Electrical results and comparison with morphological data (SEM) will be shown in detail.

## INTERCONNECT LOAD FOR METAL THICKNESS EVALUATION

A key point for a correct estimation of the metal (dielectric) thickness is the design of a proper interconnect load, design that is limited by technology issues. In copper damascene process dishing and erosion [3] don't allow to implement structures of any dimension. In particular, it is not possible to design parallel plates of large dimension without suffering of dishing effect: the result is a plate with a not constant metal thickness, thinner in the centre. The idea is to measure a capacitance which represents a good approximation of (1) by means of comb with wide fingers: the finger width can't be too narrow in order to limit fringe capacitance impact but on the other hand can't be too wide to avoid the aforementioned dishing problem. It is therefore important in order to design a proper interconnect load to know the intrinsic limits of the specific technology to characterize.

In Fig.1 the 2D cross section of the interconnect load is reported: it is based on two interleaved combs of different widths for each metal level. Comb1 (white) is the active structure used to build the capacitance *C* of interest while Comb2 (dotted) is grounded and is used to remove the fringe capacitance. By a proper load design, *C* represents a good approximation of the area capacitance: the key factor is to properly choose the width of the combs lines *w1*, *w2* and spacing *s1*. A 2D/3D field solver has to be used to select the best configuration which has the capacitance *C* closest to the ideal one in (1). The width and spacing values to use in the simulation must be tuned for a given technology.

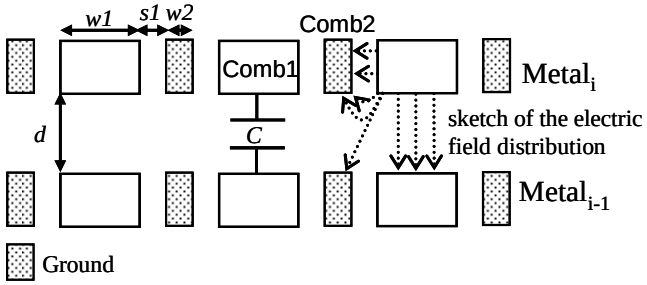



Fig. 1: 2D cross section of the proposed interconnect load for the measure of the distance *d* between two metal layers.

Table 1 summarizes a subset of the 2D simulations DOE (Design Of Experiment) performed for different widths and spacing: as can be observed, the agreement with the ideal capacitance can be very good.

**Table 1**: Subset of the 2D simulation DOE (72 experiments).

| | | | Area Capacitance | | |
|---|---|---|---|---|---|
| *w1* | *w2* | *s1* | Simulated | Ideal | % Error |
| (µm) | | | (fF/µm$^2$) | | |
| 2 | 0.18 | 0.18 | 0.1009 | 0.1019 | -0.98% |
| 2 | 0.18 | 0.36 | 0.1073 | | 5.29% |
| 2 | 0.18 | 0.54 | 0.1120 | | 9.86% |
| 2 | 0.36 | 0.18 | 0.0998 | | -2.04% |
| 2 | 0.36 | 0.36 | 0.1061 | | 4.09% |
| 2 | 0.36 | 0.54 | 0.1105 | | 8.42% |
| 2 | 0.54 | 0.18 | 0.0997 | | -2.12% |
| 2 | 0.54 | 0.36 | 0.1059 | | 3.88% |
| 2 | 0.54 | 0.54 | 0.1105 | | 8.41% |

The distance *d* can then be extracted inverting (1), where *A* is the product of the line width *w1* with the line length multiplied by the number of lines. In case of multiple dielectric layers, the dielectric constant to use in (1) is the equivalent one between the two metal levels, calculated as the dielectric constant of a parallel plate capacitor which has the same capacitance *C* and distance *d* between the plates. In formula, for two layers with dielectric constant $\varepsilon_1$ and $\varepsilon_2$ and thickness $t_1$ and $t_2$ respectively:

$$\frac{1}{C}=\frac{1}{C_1}+\frac{1}{C_2}=\frac{t_1\cdot\varepsilon_2+t_2\cdot\varepsilon_1}{\varepsilon_0\cdot A\cdot\varepsilon_1\cdot\varepsilon_2}=\frac{1}{C_{eq}}=\frac{t_1+t_2}{\varepsilon_0\cdot\varepsilon_{eq}\cdot A} \quad (3)$$

$$\varepsilon_{eq}=\frac{(t_1+t_2)\cdot\varepsilon_1\cdot\varepsilon_2}{t_1\cdot\varepsilon_2+t_2\cdot\varepsilon_1} \quad (4)$$

As can be seen in (4) the $\varepsilon_{eq}$ is a week function of the dielectric thickness, i.e. large variation in the thicknesses has a very small impact on the relative dielectric constant; this allows to use the nominal (target) values without introducing meaningful errors in the thickness estimation. In addition, due to the geometries of the interconnect loads and the selective technique used for the capacitance measurement which doesn't involve coupling capacitance for metals of the same level, even possible oxide damage around the metal in low-k process has a minimum impact.

Appling this structure to different metal layers, e.g. between $Metal_{i+1}$ - $Metal_i$, $Metal_i$ - $Metal_{i-1}$ and $Metal_{i+1}$ – $Metal_{i-1}$ it is possible to extract the $Metal_i$ thickness, as shown in Fig. 2. To describe the structure, it is better to divide it in two parts, one with all the metal levels (left side in Fig.2) and the other one (right side) with only the top and bottom metal levels. The active combs have the same width *w1* and the ground combs have the same width *w2*. In Fig.2 *s1* is the spacing between the active and grounded combs in the first part of the structure while *s2* is the spacing in the second part of the structure. The spacing *s1* and *s2* are different because it is different the electric field and so the correction that the ground combs have to make in order to obtain a capacitance *C* as close as possible to the capacitance of an ideal parallel plate. From the left side of the structure *d1* and *d2* thicknesses can be obtained, from the right side *d3* is measured: the $Metal_i$ thickness *t* can then be evaluated as *t=d3-(d1+d2)*. It is important to emphasize that the interconnect structure is only one and so all the IMD thickness measurements (*d1*, *d2*, and *d3*) can be performed on the same interconnect load. In this way, if a process variation has occurred to reduce the distance *d1*, also the distance *d3* would have been impacted by the same reduction, so the metal thickness *t* is still correctly evaluated. The implementation of a unique interconnect load is possible thanks to the capacitance measurement technique used, which can selectively extract the capacitance of interest. This is a great improvement with respect to having more than one load (one for $Metal_{i+1}$ – $Metal_i$ to obtain *d1*, one for $Metal_{i+1}$ – $Metal_{i-1}$ to obtain *d3* etc): in fact, the process variations could have a different impact for the three separated interconnect loads and so the metal thickness would have been evaluated with large uncertainty.

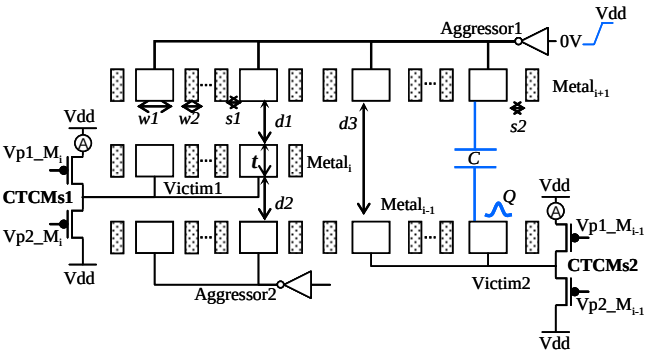



Fig. 2: Complete interconnect load scheme and relative connection to the capacitance system measurement (CTCMs). Metal thickness *t* can be obtained as *t = d3 – d2 – d1*.

Different test structures have been implemented according to Table 2. As a general guideline, *w1* must be the same for all the three metal levels while *s1* and *s2* could be different to properly correct fringe capacitance.

**Table 2**: Implemented geometries for the interconnect structures. The line length of 42µm is the same for all the structures.

| Name | *w1* | *w2* | *s1* | *s2* |
|---|---|---|---|---|
| | µm | µm | µm | µm |
| Str. 0.56um | 0.56 | 0.14 | 0.19 | 0.4 |
| Str. 1um | 1.0 | 0.14 | 0.19 | 0.4 |
| Str.1.5um | 1.5 | 0.14 | 0.19 | 0.4 |

## CTCMs TECHNIQUE APPLIED TO BACK-END CHARACTERIZATION

The accurate capacitance measurement technique used is the CrossTalk-Based Capacitance Measurements[1] (CTCMs) [4, 5, 6], an authors improvement of the well known Charge

[1] Covered by European and US (US 7352192) patent, "Method and relative test structure for measuring the coupling capacitance between two interconnect lines", by L. Bortesi et al.

Based Capacitance Measurement (CBCM) [7]. In the following the application of CTCMs to the metal thickness measurement and its basic working principle will be described while further details about CTCMs technique can be found in [4, 5, 6].

The key point of CTCMs is the capability to directly measure the capacitance between two electrodes exploiting the cross-talk effect. Fig. 2 shows two CTCMs applied to the interconnect load. For the explanation consider the right side of Fig. 2 for *d3* measurement by CTCMs2. CTCMs technique is based on two transistors of the same type (acting as independent switches) driven by two independent gate signals ($Vp1\_M_{i-1}$ and $Vp2\_M_{i-1}$) connected in series and joined to one arm of the capacitor (*Victim2*), and an aggressor signal (*Aggressor1*) connect to the other arm of the capacitor. Suppose to switch OFF the two transistors and to drive the *Aggressor1*. A voltage variation of *Aggressor1* will induce on the *Victim2* a charge *Q (Q = C*Vdd)* proportional to the coupling capacitance *C* between aggressor and victim lines and also a variation on the bias of the line. If one of the two transistors driven by $Vp1\_M_{i-1}$ and $Vp2\_M_{i-1}$ is switched ON, the additional amount of charge *Q* will be compensated by the voltage source *Vdd* and the bias on the victim line restored to *Vdd* level. If the *Aggressor1* and the gate signals are continuously driven at a frequency *Freq* as described before, the charge *Q* can be easy measured as an average current flux measured by the ammeter *A* in Fig. 2, as explained by the next formula:

$$C = \frac{Q}{V_{dd}} = \frac{I_{avg}}{V_{dd} \cdot F_{req}} \qquad (5)$$

where *Iavg* is the measured average current, *Vdd* is the aggressor line variation and *Freq* the frequency at which the operation is repeated. The timing of the driving signals involved in the measurement is detailed in [4, 5, 6]. The measured capacitance *C* can be used in equation (1) to extract the distance *d3*. In a similar way, working with *Aggressor1* and *CTCMs1* it is possible to measure *d1* and with *Aggressor2* and *CTCMs1* the *d2* thickness (the unused Aggressor and CTCMs are switched OFF). The $Metal_i$ thickness can be easily calculated as a difference of dielectrics.

## EXPERIMENTAL RESULTS AND APPLICATIONS

This paragraph will show in detail the results of the proposed technique, showing its accuracy and proposing some useful applications. Two different metal levels have been tested: electrical measures have been done for all the available dice (full wafer map) while morphological ones have been taken in a central wafer position. Data refers to a Back-End under development, and this from one side justifies the not so negligible variation across the wafer and on the other side can be considered a good check for the methodology.

*Methodology accuracy*

This paragraph will discuss the accuracy of the proposed technique by comparing the electrical measures with the morphological ones, comparing in particular the data available on the same wafer portion, i.e. the central part. Fig. 3 reports the electrical and the morphological data for the *d1*, *d2*, *d3* dielectric and metal *t* thickness for all the implemented test structures with different *w1* combs widths. Electrical data for metal 1 layer in Fig. 3 are reported in term of minimum, maximum and average values obtained over the entire wafer (wide bar) and over the central part of the wafer (narrow bar), where also SEM analysis has been performed. SEM data are indicated in the figure with dots: different sample represents different measures done at the middle and at the border of the interconnect load. Fig. 3 shows a good correlation between SEM and electrical data for each structures, in particular the structures with *w1*=1µm and *w1*=1.5µm seem a little bit better in the metal thickness *t* evaluation. As can be expected, comparison is generally better in the centre of the wafer.

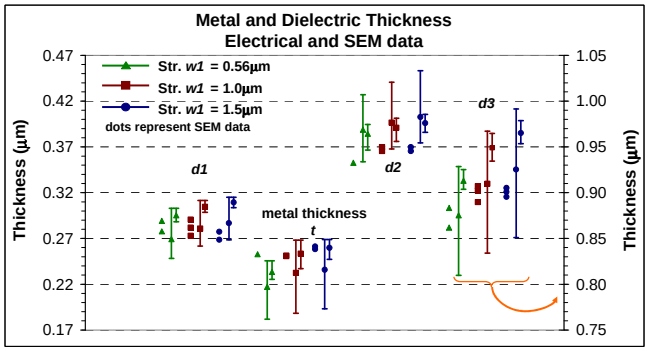



Fig. 3: Comparison between electrical and morphological (SEM) metal and dielectric thickness. Wide bars refer to measures of the entire wafer, narrow ones to measures in the central part of the wafer.

Fig. 4 reports a couple of SEM photos (referring to the left and right side of Fig. 2) for the *w1*=1.5µm structure: it can be pointed out that no dishing/erosion effects can be observed and that the morphological data for the metal thickness 0.258µm is very close to the electrical data, ranging from 0.25µm to 0.268µm in the centre of the wafer (average value 0.259µm).

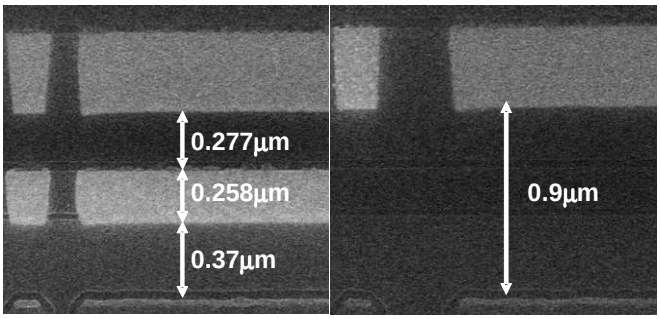



Fig. 4: SEM photos for metal and dielectric thickness for metal layer 1, *w1*=1.5µm structure.

Fig. 5a) shows the comparison for the metal2 layer thickness, where a good correlation for the *w1*=1.5µm structure but not for the *w1*=0.56µm can be observed. To better understand these discrepancies, Fig. 5b) reports the SEM photo for the *w1*=0.56µm structure: even if it is not affected by dishing it seems to suffer of notching effect which has its biggest impact on narrow line. In addition, this structure is more sensitive to metal width variation and tapering.

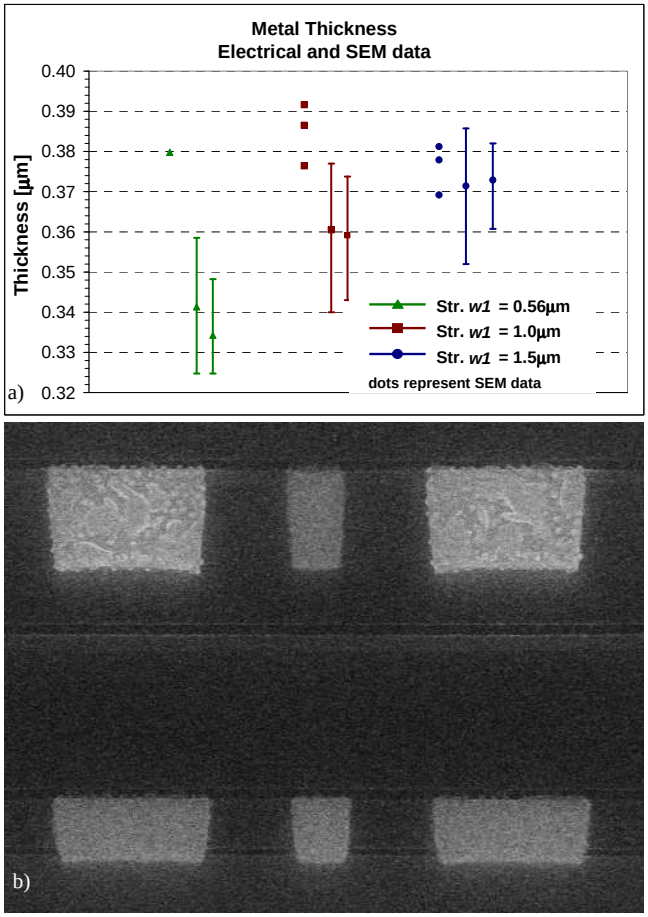



Fig. 5: a) Experimental comparison between electrical and morphological data for the metal 2 layer thickness; b) SEM photo represents the *w1*=0.56µm structure which shows the worst predictability and a notching effect for the involved metals.

For this technology, considering all the metal layers the *w1*=1.5µm structure is the best one to monitor the metal and dielectric thickness. For these structures, the predictability is on the average of 5% which can be considered a good result.

*Wafer mapping*

Due to the CTCMs sensitivity, the proposed approach can be applied to map metal (dielectric) thickness on the entire wafer to show centre-edge effects. Fig. 6a and 6b show the 3D and 2D map for the metal 1 layer thickness: it is clearly illustrated a centre-edge effect. Fig. 6c shows the metal 2 layer thickness distribution and no particular trend can be observed. In addition, by properly comparing metal and dielectric thickness with the nominal ones it could be possible to identify the most critical process steps: e.g. in a damascene process to investigate the process control over the total dielectric deposition, the dielectric etch or the metal CMP.

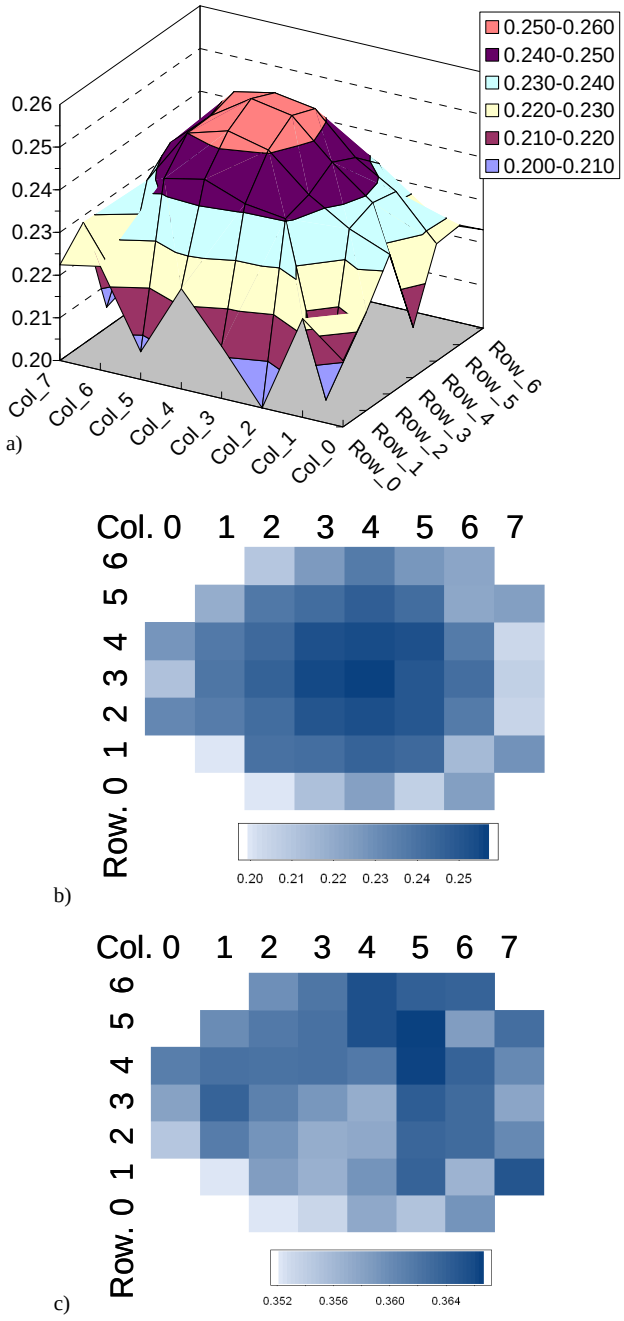



Fig. 6: Metal thickness wafer mapping: a) 3D and b) 2D view of the metal 1 layer thickness distribution over the wafer; c) 2D view of the metal 2 layer thickness distribution over the wafer.

*Metal thickness and resistance correlation*

An important application of the proposed methodology is the possibility to understand if resistance or coupling capacitance spread is due to thickness and/or width

variation. A good estimation of the metal thickness can be also useful for complex structures as RingOscillators, to correlate RC delay with interconnect parasitic variation.

Kelvin structures at minimum width and spacing have been drawn to measure the resistance of the two metal levels. Fig. 7 shows for both the metal levels the normalized resistance (respect to the minimum value) and metal thickness (respect to the maximum value) over different dice. As can be seen in Fig. 7b, the resistance variation of metal 2 layer is strictly correlated to the metal thickness variation (resistance increase due to a thickness decrease), while for the metal 1 layer there is not such a tight relation (Fig. 7a). In fact, for some points, there is a meaningful discrepancy between thickness and resistance trend possibly related to a metal width variation induced by different etching process efficiency across the wafer.

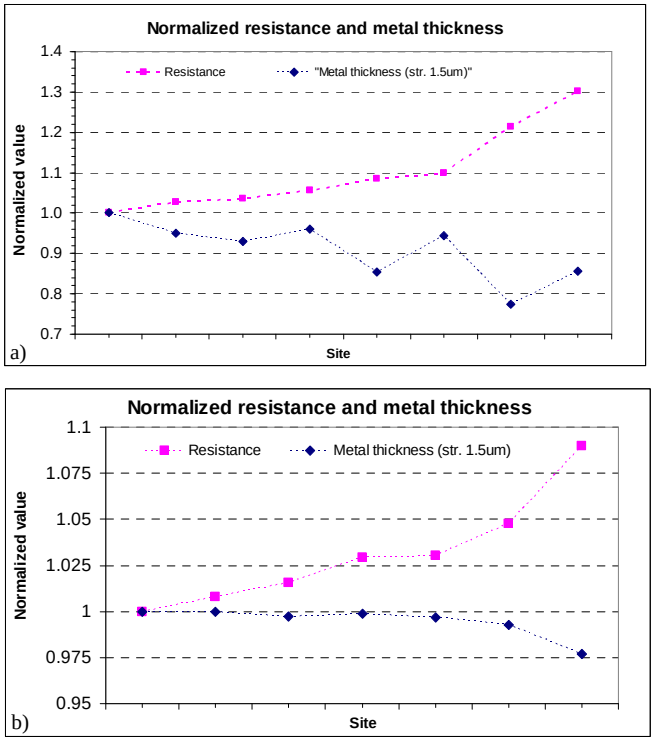



Fig. 7: Normalized resistance and normalized metal thickness a) for metal 1 layer and b) for metal 2 layer. Resistance has been normalized to the minimum value, metal thickness to the maximum value.

## CONCLUSION

A new electrical methodology for the metal and dielectric thickness characterization has been described and validated. It exploits the synergy between the CTCMs technique (a capacitance to current transducer) and the design of a proper interconnect load to directly and selectively measure the involved capacitances to evaluate all the thicknesses of interest. The proposed approach can be easily integrated in a parametric test due to its simple implementation and its low silicon area consumption, compatible with saw-line insertion. Thanks to its resolution it allows to detect small variations from the BEOL target. Its absolute accuracy below 5% is related to the optimization of the interconnect structure, the correct estimation of the equivalent dielectric constant and the approximation between rectangular and true line shape.

## ACKNOWLEDGEMENTS

The authors whish to thank STM Agrate pilot line for supplying samples for comparison purposes.

## REFERENCES

[1] H. Kinzelbach, “Statistical variations of interconnect parasitics: extraction and circuit simulations”, *Proceeding of 10th Workshop on Signal Propagation on Interconnect*, 2006, pp 33-36.

[2] V. Nguyen Hoang, A. Kumar, P. Christie;., “The impact of Back-End-Of-Line process variations on critical path timing”, *Proceedings of the 7th International Symposium on Quality Electronic Design*, 2006.

[3] A. Toffoli et all., “Test structures for characterizing metal thickness in damascene CMP technology”, *ICMTS 2008*, pp 210-213.

[4] A. Bogliolo, F. Vaira, L. Vendrame, L. Bortesi, “A new test structure for measuring on-chip cross-coupling capacitances”, *Proc. IEEE Signal Propagation Interconnect Conf*., 2004, pp. 165–166.

[5] L. Vendrame, L. Bortesi, A. Bogliolo, “New applications of Cross-Talk-Based capacitance measurements”, ICMTS 2005. *Proceedings of the 2005 International Conference on Microelectronic Test Structures*, 2005, p. 257-261.

[6] L. Vendrame, L. Bortesi, F. Cattane, A. Bogliolo, “Cross-Talk-Based capacitance measurements: theory and applications”, *IEEE Transactions on Semiconductor Manufacturing*, Vol.19, No.1, February 2006.

[7] J. C. Chen, B. W. McGaughy, D. Sylvester, C. Hu, “An on-chip, attofarad interconnect charge-based capacitance measurement (CBCM) technique”, *Proceedings of IEEE IEDM*, 1996, pp.69-72.

# A Test Structure for Assessing Individual Contact Resistance

Frank Liu
IBM Austin Research Lab
Austin, Texas 78758, U.S.A.
Email: frankliu@us.ibm.com

Kanak Agarwal
IBM Austin Research Lab
Austin, Texas 78758, U.S.A.
Email: kba@us.ibm.com

***Abstract*—Accurate measurement of contact resistance is crucial for advanced nanometer CMOS processes. An equally important requirement is to measure contact resistances in the same micro-environment as the *device-under-test* (DUT) will be used in real designs. With complicated interactions among various layout shapes in nanometer CMOS processes, test structures with adequate scalability is needed. In this paper we present a scalable contact resistance measurement structure, which can accommodate tens of thousands of DUTs. The measurement results from a 65nm CMOS technology are also presented.**

## I. Introduction

As the scaling of nanometer CMOS processes continues, the cross-section area of the contact plugs has to be scaled accordingly, while their height remains relatively flat from one generation to another. As a consequence, the contact resistance has increased drastically. Accurate assessment of the contact resistance and its distribution are critical for process characterization and yield learning.

An equally important requirement for contact resistance measurement is that the resistances should be measured in the same micro-environment as the *device-under-test* (DUT) will be eventually used in real designs. Typical micro-environment includes the relative positions of the contacts in the immediate neighborhood, as well as the contact overlay with respect to M1 or polysilicon shapes. The reason is that as the scaling is further pushed, many systematic components of the variability are highly influenced by the detailed layout of the structure. An example comes from the fact that the overall contact resistance is a function of the M1 overlay relative to the contact. With the wide use of resolution enhancement techniques (RET), the exact M1 overlay over the contact could be strongly influenced by other M1 shapes in the immediate neighborhood. In order to make an accurate assessment of resistance variability, it is highly desirable to maintain the local M1 environment.

Traditionally the measurement of the contact resistance has been carried out by implementing the well-known contact chains [3], in which hundreds to thousands of contact links are connected in a series fashion. The average resistance of the contacts can be calculated by dividing the overall resistance of the whole chain with the number of the contact links. However, the statistical information (i.e., mean resistance) provided by the chain is highly limited. The chain structure is further hampered by the fact that it lacks the capacity and flexibility to accommodate wide spectrum of multiple design patterns with different micro-environments. In recent years, there have been several proposals of individually addressable contact resistance measurement structures. In the structures proposed in [2] and [4], rows of individually addressable contact chains are formed. Each link in the chain can then be measured by individually activating the intersecting column. A more recent extension was proposed in [1]. Although quite elegant and providing much needed flexibility, the capacity of these structure is somehow limited by the minimal resolution since each row of DUTs is activated, those DUTs effectively form a voltage divider.

In this paper, we present a contact resistance measurement structure, in which each DUT can be individually activate and measured. The structure can easily scaled thus provide desirable capacity and flexibility. The rest of the paper is organized as follows: in Section II we present the details of the test structure design; in Section III we present the measurement results and some preliminary analysis, followed by conclusions in Section IV.

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

## II. Description of Test Structure

The contact resistance is measured by applying the basic four-point Kelvin configuration. Whenever a particular DUT is activated, a given current source is diverted to the DUT and the voltage across the DUT is measured. The core of the test structure is an array which consists of $119 \times 240$ cells, each of which contains a testable contact (or contact clusters such as double or triple contact arrays). Note that the choices of the number of rows and columns are purely arbitrary. The proposed structure can be expanded or reduced with relative ease. A simplified version of a $4 \times 4$ array is shown in Fig. 1. For each column of the core array, there are two vertical

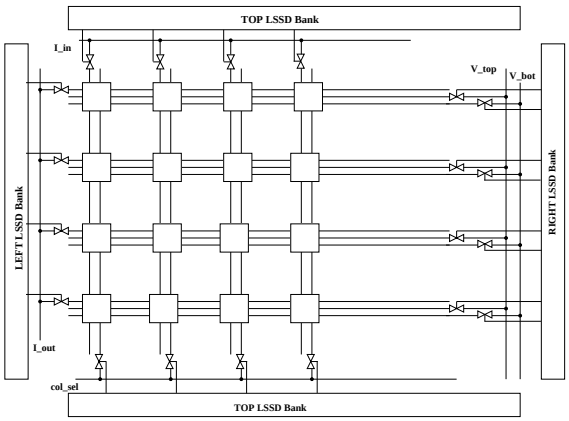



Fig. 1. The simplified schematic diagram of the contact measurement array. Note only a $4 \times 4$ array is shown.

global wires (which we refer to as the "buses"): $I_{in}$ and the column enable signal. For each row of the array, there are three horizontal buses: $I_{out}$, $V_{top}$ and $V_{bottom}$. The purpose of the $I_{in}$ and $I_{out}$ wire is to provide the necessary current to the DUT, while $V_{top}$ and $V_{bottom}$ are high impedance measurement points for voltage measurement. The column enable signal is needed to make sure that there is no internal current loops. The detailed schematics within each cell is show in Fig. 2. The $V_{top}$ and $V_{bottom}$ wire if each row are connected to the global $V_{top}$ and $V_{bottom}$ via switches on the right side of the core macro. Similarly, the $I_{out}$ wire of each row is connected to the global $I_{out}$ wire via pass-gate switches on the left side of the core macro. By properly enable/disable those switches, we can make sure that only the selected row is activated and measured. Same principal applies to the $I_{in}$ and column enable signal of each column, except that the switches reside on the top and bottom of the core array. The resistance of each DUT can be easily calculated by applying the Ohm's

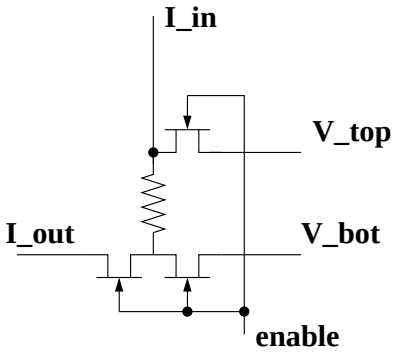



Fig. 2. Schematic diagram of the DUT cell.

Law on the current and the voltage across. This two-level addressing scheme achieves the desirable flexibility. To further illustrate the nature of the core array, the schematic of a $2 \times 2$ example is shown in Fig. 3. In the figure, the circled DUT is selected and the direction of the current is indicated with arrows. However, due to the unavoidable leakage current in the nanometer CMOS devices, there are also leakage paths from unselected DUTs, as shown with dashed lines in the same figure. Obviously the leakage will have negative effect on the measurement data. To address this issue, the column select signal of each column is implemented as one-hot selection: it is either connected to a logic 1 or a clamping voltage, which can be set further below 0. This technique provides significant reduction on the leakage current.

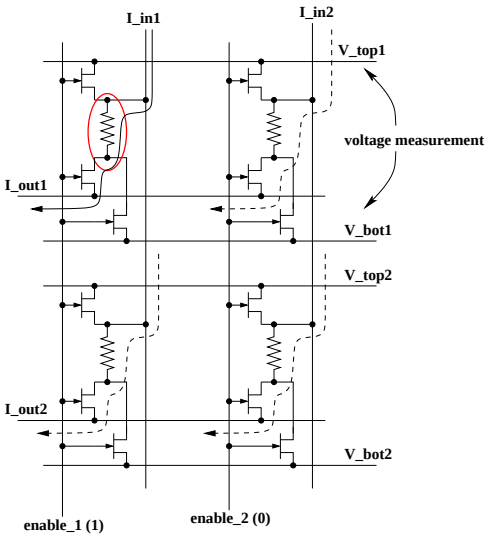



Fig. 3. Schematic diagram of $2 \times 2$ array.

As indicated in the previous discussion, there are pass-gate switches on all four sides of the core array. These switches are controlled by *Level Sensitive Scan*

*Design* (LSSD) latch banks on all four sides of the core array. To minimize the number of required IO pins, they are chained together to form a single scan-chain. It is also possible to implement additional logic so that the row/column selection signals are generated in a sequential fashion so that each row (or each column) or even the whole core array is activated sequentially.

## III. Measurement Results and Discussion

The test macro has been implemented and manufactured in a $65nm$ bulk CMOS technology. Several different contact configurations were implemented, e.g., the relative location of the contact within the diffusion region etc. The top-view of the macro is shown in Fig. 4. Note that we didn't include the pad-set in the picture. One can clearly see the peripheral circuits on the four sides of the core array. The overall size of the macro is approximately $300 \times 600\mu m$.

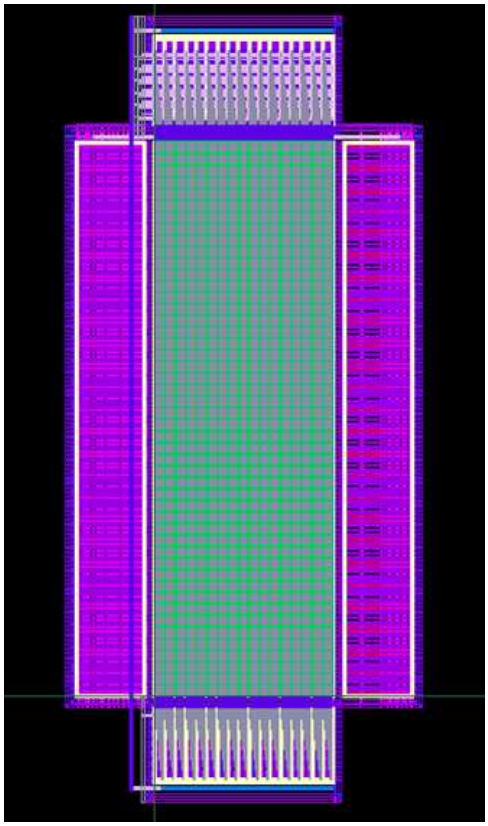


Fig. 4. Layout view of the macro. The size of the macro is approximately $300 \times 600\mu m$.

The measurement of the macro is conducted at wafer level by using a Keithley 4200 semiconductor analyzer. To see the die level distributions of the contact resistances, the histograms of two layout configuration types are shown in Fig. 5. At this particular die, there

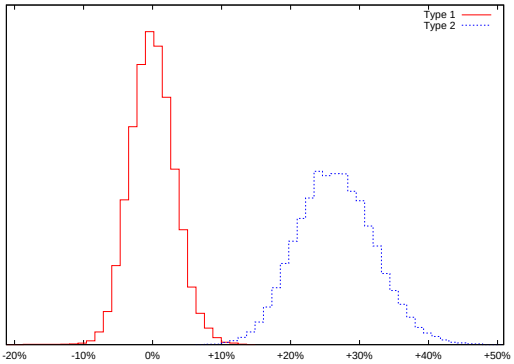



Fig. 5. Die level resistance distribution of two layout configuration types. Note that the resistance values are scaled to reflect the percentage difference between two configuration types.

is about $25\%$ difference between the mean resistance values of these two types of configurations. The standard derivation of the Type 2 configuration is slightly larger than that of Type 1. It is also quite clear that the resistance values of both types are very close to Gaussian distribution. To confirm this, the qqplot of the two sets of data is shown in Fig. 6. It is clear that the distributions are very close to Gaussian distribution up to $\pm 3\sigma$.

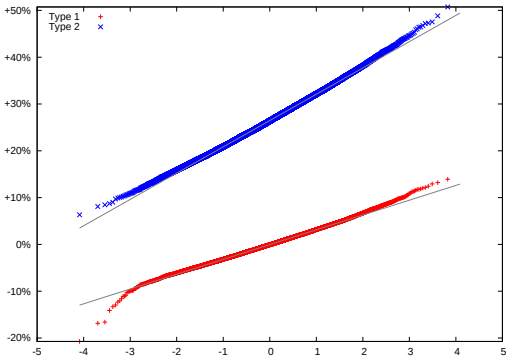



Fig. 6. Qqplot of the resistance values of two layout configurations.

However, the full wafer test data shows a different trend. Fig. 7 shows the histogram of the resistance distributions of Type 1 layout configuration. In order to emphasize the "tail" of the wafer-level data, the y-axis of the plot is in log-scale. The long tail of the data is quite obvious. Fig. 8 shows the qqplot for the same set of data.

To investigate the disparity between the die-level data distribution and the wafer level trend, we further conducted data analysis for each die. The mean values at each die along with their standard derivations are plotted in Fig. 9 using error bars. It is clear that although majority of the dies have relatively comparable mean

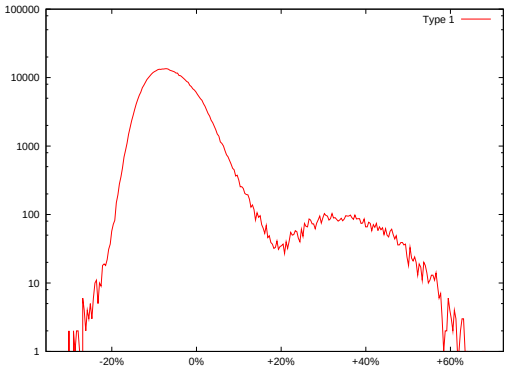



Fig. 7. Histogram of Type 1 layout configuration contacts at full wafer level. Note y-axis is in log-scale.

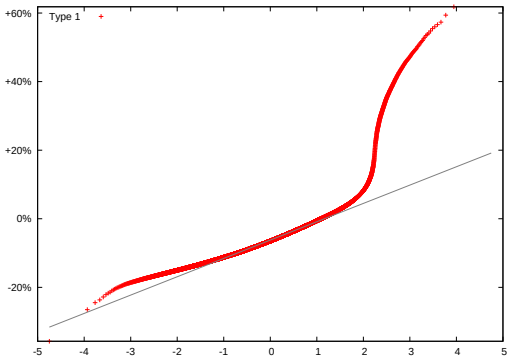



Fig. 8. Qqplot of contacts of Type 1 layout configuration at full wafer level.

and standard derivation values, the measurement results from a few dies are clearly shifted from the rest. In particular the results from one die are $30\%$ larger than the rest of wafer, which explains the existence of the long tail in the wafer level data. We further conducted

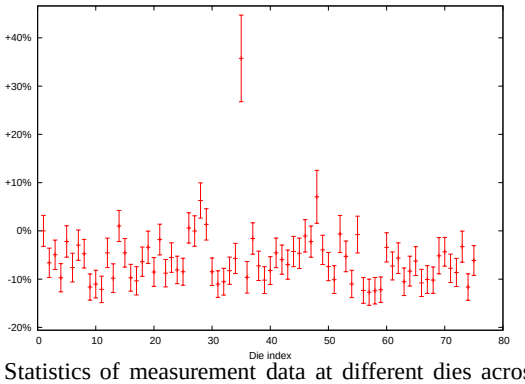



Fig. 9. Statistics of measurement data at different dies across the wafer. The middle points in the error bar indicate the mean values, while two end points represent the $+1\sigma$ and $-1\sigma$ quantile points, respectively.

statistical analysis on the Type 1 data on this particular die. The qqplot of the resistance values is shown in Fig. 10. Although the data is slightly off compared to an ideal Gaussian distribution, overall the Gausianality of the data is still quite good. This is an indication that there is no systematic issues with the data. Given the fact that there is only one outlier among dozens of dies, we have to treat it as a statistical abnormality. After discarding the only outlier, we can see from Fig. 9

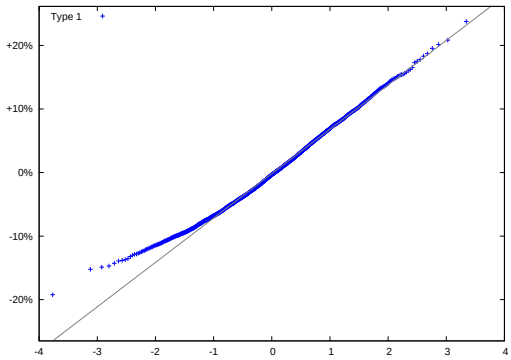



Fig. 10. Statistics of measurement data on the outlying die.

that when compared to the intra-die portion, the inter-die portion of the resistance distribution is still quite significant. Many non-overlapping error bars at different die locations indicate that more careful analysis on how the variability distributes is needed.

## IV. Conclusion

In this paper, we present a scalable contact resistance measurement structure. The structure has the capability of individually addressing tens of thousands of DUTs. Measurement results from a $65nm$ CMOS technology indicates that the structure functions properly. Full wafer measurement results show that inter-die component of the resistance variability is comparable to the intra-die component.

## References

[1] T. Hamamoto, T. Ozaki, M. Aoki, and Y. Ishibashi. Measurement of contact resistance distribution using a 4k-contacts array. *IEEE Transactions on Semiconductor Manufacturing*, 9(1):9–14, Feb 1996.

[2] N. Kasai, H. Koga, and Y. Takaishi. Contact resistance measurement of a 130-nm-diameter poly-Si plug on a lightly doped single diffusion region in giga-bit DRAMs. In *IEEE International Conference on Microelectronic Test Structures*, pages 177–181, 2001.

[3] H. Sayah and M. Buehler. Comb/serpentine/cross-bridge test structure for fabrication process evaluation. In *IEEE International Conference on Microelectronic Test Structures*, pages 23–28, 1988.

[4] H. Sayah and M. Buehler. Linewidth and step resistance distribution measurement using an addressable array. In *IEEE International Conference on Microelectronic Test Structures*, pages 87–92, 1990.

# Fast Embedded Characterization of FEOL Variations in MOS Devices

F.Rigaud[1&2], J.M. Portal[1], P. Dreux[2], J.Vast[2], H. Aziza[1], G.Bas[2]

[1]IM2NP - Laboratoire Matériaux et Microélectronique de Provence, UMR CNRS 6242, IMT-Technopôle de Château Gombert, F-13451 Marseille Cedex 20, France
[2]ST Microelectronics, ZI Rousset, BP2, F-13106 Rousset, France
*fabrice.rigaud@l2mp.fr*

## ABSTRACT

The objective of this paper is to present a test chip based on embedded Ring Oscillators (RO) measurement with its associated extraction algorithm to characterize length and width variations and to discriminate them from others FEOL variations. A brief overview of the structure, designed in a ST-Microelectronics 90nm technology, is given with emphasis on the ROs geometry with their biasing conditions and the measurement circuit. Comparison of simulated values versus estimated ones is given and confirms the ability of the structure to characterize FEOL variations. MOS width and length are well estimated regardless the others FEOL deviations that can be also detected.

## INTRODUCTION

Today, complex circuits are manufactured using advanced technologies, which are more prone to large Process Variation (PV). In order to maintain high yield, process drift monitoring is one of the major issue of semiconductor manufacturing. Some structures are already used to evaluate transistor performances [1] or to extract threshold voltage in many layout configurations [2]. Another way consists to monitor intra-die PV inserting Ring Oscillators (ROs) across product [3].

The main objective of this paper is to propose a test chip and its associated test data analysis methodology to characterize front-end variations, i.e. transistor geometry and doping density. It is of prime importance to characterize those variations, because of their straight impact on the electrical performances of the devices (current and threshold voltages). At circuit level, these variations can result in yield loss due to excessive delay overhead. Knowing that front-end variations impact the delay, the proposed test-chip is based on ring oscillator.

The first section of this paper describes the test chip architecture based on a set of ring oscillators surrounded by an embedded measurement circuit. The second section presents the frequency measurement analysis method which discriminates geometric variations from others process variations and estimates transistor width and length. This method is validated through simulation results presented in the last section.

## TEST STRUCTURE DESIGN

The core of the test structure is based on a set of three ring oscillators (RO). In order to characterize the transistor width and length variations and to discriminate them from others front-end process variations, the ROs devices are dimensioned as summarized Table I.

TABLE I
RO WIDTH AND LENGTH SELECTION

| | NMOS | | PMOS | |
|---|---|---|---|---|
| RO identification | Width | Length | Width | Length |
| *RO1* | $W_{MIN}$ | $L_{MIN}$ | $k.W_{MIN}$ | $L_{MIN}$ |
| *RO2* | $W_{MAX}$ | $L_{MIN}$ | $k.W_{MAX}$ | $L_{MIN}$ |
| *RO3* | $W_{MIN}$ | $L_{MAX}$ | $k.W_{MIN}$ | $L_{MAX}$ |

The minimal geometries are chosen in order to amplify the effect of any dimensional variations. Moreover, the ROs are tested using different bulk bias conditions to discriminate the impact of other front-end variations from dimensional variations. The ROs are surrounded by a triple-well to isolate the ROs from the rest of the test structure. The different bulk bias conditions are given Table II.

TABLE II
BULK BIAS CONDITIONS

| | Bias 1 | Bias 2 | Bias 3 | Bias 4 | Bias 5 |
|---|---|---|---|---|---|
| NMOS $V_{SUB}$ (V) | gnd | $V_{DD}$ | $V_{DD}/2$ | $V_{DD}$ | gnd |
| PMOS $V_{SUB}$ (V) | gnd | $V_{DD}$ | $V_{DD}/2$ | gnd | $V_{DD}$ |

The substrate bias condition allows changing the threshold voltages of the devices in order to better control the impact of the others Front-End parameters (doping concentration

and profile, thickness etc…), called FEN for the NMOS and respectively FEP for the PMOS.

TABLE III
RO FREQUENCY (MHz) VS BIAS CONDITIONS

| | Bias 1 | Bias 2 | Bias 3 | Bias 4 | Bias 5 |
|---|---|---|---|---|---|
| *RO1* | 326 | 278 | 351 | 354 | 274 |
| *RO2* | 363 | 315 | 396 | 398 | 312 |
| *RO3* | 246 | 210 | 265 | 268 | 207 |

Table III shows the RO frequency in MHz for all bulk biasing conditions. As expected, when reverse biasing is applied on the bulk, the threshold voltage decreases and thus the frequency increases. Indeed, the standard biasing (bias 5) presents the lower frequency whatever the RO considered, and the complete reverse biasing (bias 4) presents the higher ones.

An embedded measurement structure is designed to extract RO frequency on chip. The layout of the structure is given Fig. 1. This structure is fully digital and has only one master clock input (MS_CLK) and one output (DOUT). Its architecture (Fig. 2) consists of:

- One state machine.
- A time reference given by a 12-bit counter so called time base counter (TBC).
- One 16-bit counter associated with each RO, so called RO counter (ROC).
- A multiplexer to select the RO counter.

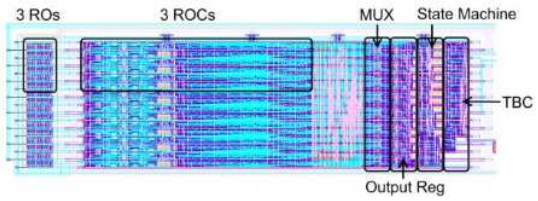



Fig. 1. Test structure layout

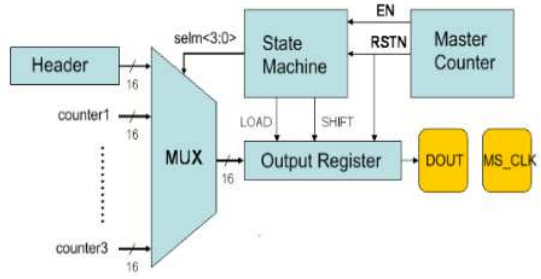



Fig. 2. Functional view of test structure

- One data output 16-bit register with parallel inputs and serial output.

When the MS_CLK is activated (30 MHz), the TBC starts to count in order to define a time window. Half of this window is dedicated to ROC counting (ROs oscillate and control the clock input of their associated ROC). The other half is devoted to the extraction of each ROC's content through the multiplexer in the output data register. Then the data are shifted on the DOUT pin. It is important to note that the state machine allows controlling the entire process and generating a header code primary to any data on the output. In that way, the three RO frequency data are given by a 64-bit word (16 bits of header code plus 16 bits for each RO).

$$T_w = N_{clk} * T_{clk} = 1021 * 33.33 \times 10^{-9} = 34.033\ \mu s \quad (1)$$

$$F_{osc} = \frac{N}{T_w} = \frac{N}{34.033 \times 10^{-6}} \quad (2)$$

The frequency calculation is given (1), where $T_W$ is the half time window (counting time), $N_{clk}$ the number of clock periods included in the $T_W$, Tclk the period of the clock and (2) where $F_{osc}$ is the ROs' oscillation frequency and N the number of oscillating periods during $T_W$.

## FREQUENCY ANALYSIS METHOD

This section is dedicated to the analysis method used to characterize length and width variations and to discriminate them from others front-end variations from the frequency measurement. This method is based on a DOE approach to define a set of variations (dL, dW, dFEN and dFEP) for the three ROs. Then simulations are performed for each RO at each biasing condition. The different variations given by the DOE are directly set in the transistor model. The frequency extraction for all the experiments allows obtaining RO frequency as a function of dL (length variation), dW (width variation), dFEN (others front-end variations for the NMOS) and dFEP (others front-end variations for the PMOS). Equation 3 gives the RO frequency as a function of xi, with xi representing the variation of the parameter i normalized versus the center value and the maximal deviation.

$$f(x) = b_0 + \sum_{i=0}^{3} b_i \times x_i + \sum_{i=0}^{3} b_{ii} \times {x_i}^2 + \sum_{i,j=0}^{3} b_{ij} \times x_i \times x_j \quad (3)$$

Following this approach, a set of 15 equations (3 ROs with 5 biasing conditions) is obtained. The corner stone of this approach is to solve this system to obtain the characterized value of the dL, dW, dFEN and dFEP variations with the 15 frequency values extracted on the structure at the end of the

test flow. To do so, a Newton-Raphson (N-R) algorithm with a variable convergence step [4] is applied on the system. The initial values used to solve the system are the central values of the technology considered. The algorithm consists in reducing the difference between the frequency estimated with the initial value and the frequency measured. Then, the system F(x) is transform to the system G(x) in order to use N-R algorithm (4).

$$\begin{bmatrix} F_{osc1_{bias1}} = f_1(x_{ini}) \\ F_{osc2_{bias1}} = f_2(x_{ini}) \\ \dots \\ F_{osc3_{bias5}} = f_{15}(x_{ini}) \end{bmatrix} \Leftrightarrow \begin{bmatrix} 0 = g_1(x_{ini}) \\ 0 = g_2(x_{ini}) \\ \dots \\ 0 = g_{15}(x_{ini}) \end{bmatrix} \quad (4)$$

Each convergence step is computed as described (5), (6), (7) where G is the system of equation to solve, J the Jacobian matrix, $\partial x$ the step value, $x_{old}$ and $x_{new}$ respectively the initial value of current iteration.

$$G(x) = 0 \quad (5)$$

$$\partial x = -J^{-1} \times G \quad (6)$$

$$x_{new} = x_{old} + \partial x \quad (7)$$

However the N-R algorithm does not always converge especially if the initial value is far from the solution. That is why a globally convergent method using the N-R algorithm has been used. This method consist to evaluate the convergence taking the entire value of $\partial x$, then if convergence is acceptable, the step is kept, but if not the step is multiplied by a coefficient α. Furthermore if we try to solve the system with the 15 equations, the algorithm may not converge therefore we solve multiple sub-systems of 6 equations and compute the average value of the solutions.

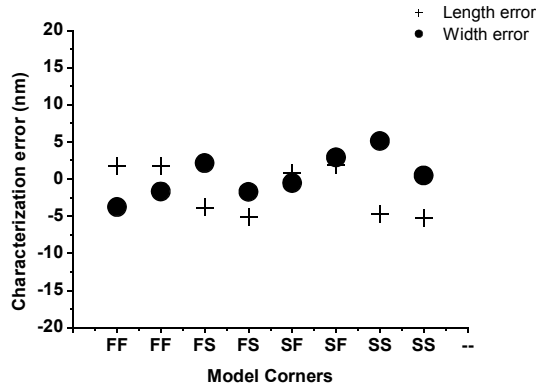



Fig. 3. Plot of the dW and dL errors between the simulated and the characterized values for all the process corners

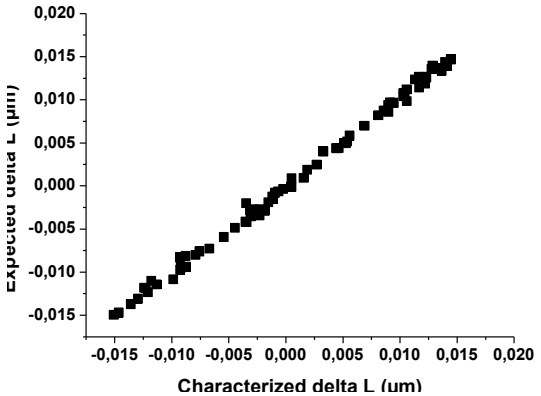



Fig. 4. Correlation between simulated and characterized Length variations

## APPROACH VALIDATION

### A. Results

The validation of the structure and of the analysis method efficiency is performed with a set of two experiments.
In the first experiment, the structure is simulated using the 4 process corners (Fast-Fast, Fast-Slow, Slow-Fast and Slow-Slow) without L and W corners, in order to evaluate the characterization capability of the method versus extreme FEN and FEP variations. To do so, two set of dL and dW variations are introduced and characterized for each corners. The results confirm the good ability of the approach to characterize dL and dW even with extreme variation of the others front-end parameters dFEN and dFEP. Fig. 3 shows that both dL and dW are characterized with a maximal error of 5nm for a 90nm technology. In the second experiment, a Monte-Carlo simulation on dL, dW, dFEN and dFEP with 100 draws is applied on the structure. Fig. 4 and 5 show the good correlation between the characterized dL, dW and the simulated ones whatever the other front-end variation. From these results, it clearly appears that using this approach, one has the capacity to characterize length and width variations and to discriminate them from others front-end variations (Fig. 6).

### B. Discussion

Results of the structure can be discussed considering BEOL variations or accuracy of the time based counter which suffer the same variations than the ROs.

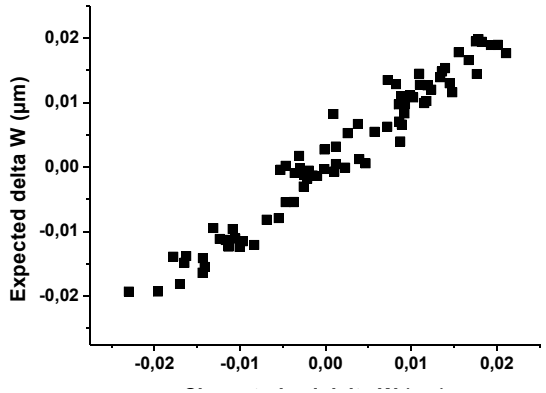



Fig. 5. Correlation between simulated and characterized Width variations

All reported simulations have been performed using netlist with typical parasitic extraction (RCTYP). So in order to evaluate the impact of BEOL variations, others simulations have been run with maximum parasitic extraction (RCMAX). These simulations show, for this extreme condition, that:

- The method is no more usable with the 3 ROs.
- Taking oscillators with a lower W/L ratio, the oscillation frequency is less impacted by the BEOL variations. Indeed, the time constant introduced by the BEOL become insignificant compared to the transistor resistivity, then the method is less accurate but usable.

Furthermore, Fig. 1 illustrates that oscillators with a lower W/L ratio are available on the structure.

Concerning the time based counter which is used for giving the ROC counting duration, simulations have been run to evaluate the impact of FEOL and BEOL variations. On one hand, BEOL variations, even for RCMAX extraction, have no impact on the counting time. On the other hand, FEOL variations change the counting time duration. Indeed, running simulations with a worst case FEOL variation (SF corner), it appears that $T_w$ increases from 34.033µs to 34.099µs. Using (2), a ring oscillating at 1GHz will be estimated at 1.001929GHz that is 0.194% of error, and a 100MHz oscillator will be measured at 100.166MHz that is 0.176% of error. In other word, because $T_W$ is very long, the short variation of this duration does not affect the oscillation frequency estimation.

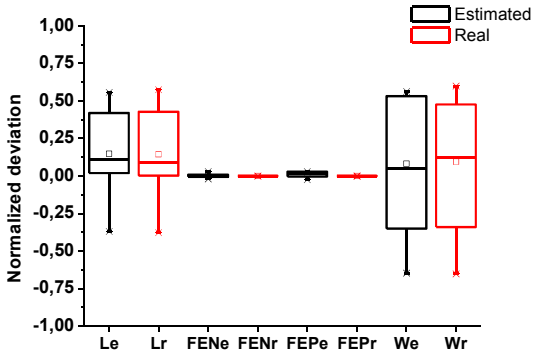



Fig. 6. Length and width variations are discriminate from other FEOL variations

## CONCLUSION

In this paper, a test chip based on embedded RO measurements and extraction algorithm is presented in order to characterize length and width variations and to discriminate them from other FEOL variations for NMOS and PMOS devices. This structure is limited to three RO with various geometries and need only one input and one output, consequently it could be embedded in a product.

The extraction of the frequency measurement is fully digital and can be performed on low cost digital tester. The test-chip is designed on a ST-Microelectronics 90nm technology, and the first experiments show that the length and width values can be characterized with a maximal error of 5nm even for FEOL corner cases, and that solution exists in case of consequent BEOL variations. Furthermore, it is shown that both FEOL and BEOL variations will not affect the accuracy of the frequency measurement which validates successfully the approach.

## REFERENCES

[1] Saxena, S et al., "Variation in Transistor Performance and Leakage in Nanometer-Scale Technologies", Electron Devices IEEE Transactions on Volume 55, Issue 1, pp.131–144, Jan. 2008.

[2] K. Agarwal, S, Nassif, "Characterizing Process Variation in Nanometer CMOS", Design Automation Conference, pp.396-399, June 2007.

[3] M. Nourani, et al, "Testing on-die process variation in nanometer VLSI", IEEE Design & Test of Computers, Vol. 23, N°6, pp.438-451,November-December 2006.

[4] Broyden, "Mathematics of Computation", C.G. 1965, vol. 19, pp. 577–593.

# Benefit of Direct Charge Measurement (DCM) on Interconnect Capacitance Measurement

Yasuhiro Miyake
Masaharu Goto

Agilent Technologies International Japan, Ltd.
9-1 Takakura-cho, Hachioji-shi, Tokyo, 192-0033, Japan

## ABSTRACT

This paper discusses application of direct charge measurement (DCM) on characterizing on-chip interconnect capacitance. Measurement equipment and techniques are leveraged from Flat Panel Display testing. On-chip active device is not an essential necessity for DCM test structure and it is easy to implement parallel measurements. Femto-Farad measurement sensitivity is achieved without having on-chip active device. Measurement results of silicon and glass substrates, including parallel measurements, are presented.

## INTRODUCTION

As Ultra Large Scale Integration of silicon devices continues to evolve, the interconnect capacitances have become a significant factor in determining the device speed. Hence, accurate characterization and modeling of interconnect capacitances become critical to a successful circuit design. In recent years, low-k inter-layer dielectrics are introduced to reduce interconnect capacitances. Low-k materials are inherently less mechanically and chemically robust. In order to ensure target performance and reliability, many measurements have to be made during the course of process development and production ramp up.

Charge-Based Capacitance Measurement (CBCM) has been widely used to characterize on-chip interconnect capacitances in recent years [1] [2] [3] [4] [5] [6]. CBCM has very good resolution that allows measurements of capacitances in the Femto-Farad range. Nevertheless, a CBCM test structure requires on-chip active device, it can be used only after process integration. Conventional LCR meter measurement does not require on-chip active device and has good measurement resolution. However, parallel measurements are difficult due to signal coupling. This limits measurement throughput and number of measurements realistically made in an acceptable timeframe. As further development of lower-k material continues, efficient capacitance evaluation method that can be used from early process development stage with high throughput is desirable.

Direct charge measurement (DCM) has not been seriously considered as viable interconnect capacitance measurement option. DCM theory is simple, however, it needs sophisticated measurement instrumentation including high-precision low-noise integrator, reset circuitry and leakage compensator [7]. Absence of off-the-shelf DCM equipment made it difficult to use this technique. In this paper, we leverage Agilent88000 HS100 flat panel display (FPD) tester on interconnect capacitance measurement. DCM has been commonly used in FPD testing [20] [21] [22], and also used in MEMS device testing [23]. Femto-Farad measurement resolution is achieved without on-chip active device. To improve measurement throughput, parallel measurements are done without adding significant measurement error.

## DIRECT CHARGE MEASUREMENT (DCM) THEORY

We will review basic theory of DCM. As shown in Figure 1, DCM instrumentation consists of an op-amp integrator and a step voltage generator. When idle, integrator reset switch is turned on in order to discharge integration capacitor Cr. At measurement, reset switch is released and then step voltage is applied on one of the target capacitor electrode. Integrator input is connected to the other capacitor electrode for reading out injected charge. With known step voltage Vstep and the integration capacitor Cr, by measuring the op-amp output voltage step Vdiff, we can calculate target capacitance C1 as follows.

$$C1 = - Vdiff / Vstep*Cr \quad \cdots\cdots (1)$$

Here, two factors, parasitic capacitance and leakage current on a sense line, can cause measurement errors.

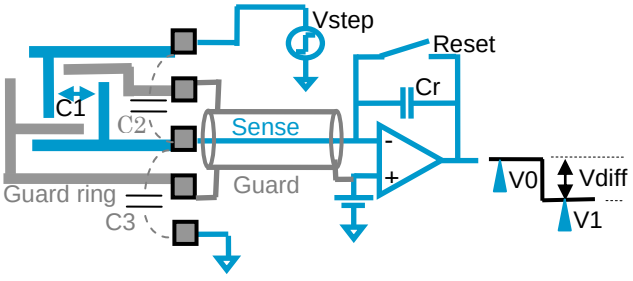



Figure1. Direct Charge Measurement Theory

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

### Parasitic Capacitance Cancellation

Effects of parasitic capacitances are essential error factors that any measurement methods, such as CBCM and LCR meter measurement, commonly have. Figure 2 shows equivalent circuit of DCM with parasitic devices C2, C3, R2 and R3.

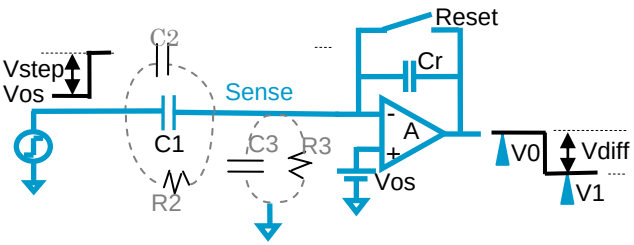



Figure 2 DCM equivalent circuit

Assuming Vos=0V and leakage resistance R2 and R3 are high enough to be ignored, we have simultaneous equations (2) to (6).

$$Q1+Q2 = Q3 + Qr \quad (2)$$

$$Q3 = C3*Vi \quad (3)$$

$$Q1+Q2 = (C1+C2)*(Vstep - Vi) \quad (4)$$

$$Qr = Cr*(Vi - Vdiff) \quad (5)$$

$$Vdiff = - A*Vi \quad (6)$$

Where Q1+Q2, Q3 and Qr are charge values of C1+C2, C3 and Cr respectively, A is a gain of the op-amp, Vi is a sense line voltage. We can solve these equations as follows. This is a strict version of equation (1).

$$C1+C2=-Vdiff/Vstep*(Cr*(1+A)+(C1+C2+C3))/A \quad (7)$$

With a high speed op-amp, the gain A is normally >100dB or $>10^5$ for low frequency and >40dB or >100 even within high-speed measurement bandwidth which means A>>1. We choose Cr to be few Pico-Farads. If C1+C2+C3 are reasonably small, the measurement error caused by C1+C2+C3 can be ignored. Then, (7) is simplified as (8)

$$C1+C2 = -Vdiff/Vstep *Cr \quad (8)$$

Offset capacitance C2 should be minimized by careful guard ring design. And we do offset cancel measurement using zero-farad reference capacitor to subtract obtained C2. This offset cancellation technique is common for any measurement methods, hence, we do not explain in detail here.

### Leakage Current Canncellation

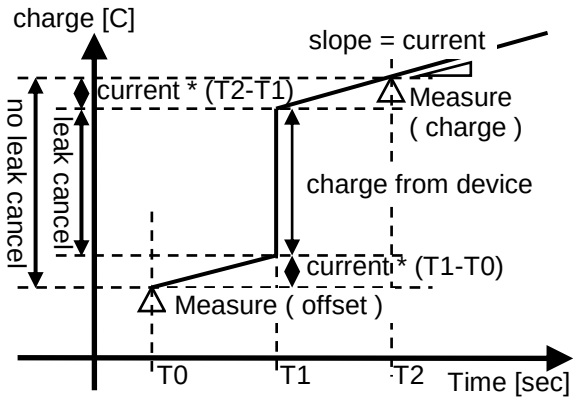



Figure3. Effect of leakage current in capacitance measurement

Effect of leakage current is depicted in Figure3. At T0, we measure the integrator output as an offset. At T1, we apply a step pulse to inject charge through the measured capacitance C1. At T2, we measure again the integrator output. Slopes of figure 3 are caused by leakage current through R2 and R3. These slopes cause measurement error. From the equivalent circuit shown in Figure 2 with non-zero Vos, leakage current during T0 to T1 is

$$i01 = - Vos/R3 \quad (9)$$

and leakage current during T0 to T2 is

$$i12 = Vstep/R2 - Vos/R3 \quad (10)$$

In order to accurately cancel error from the leakage current, we need to measure slopes at T0 and T2 separately and compensate with different slope factors. If we assume the leakage currents are stable during T0-T1 and T1-T2, those slopes are constant. Therefore, the error from the leakage current can be eliminated by measuring the slopes (dV/dt) at T0 and T2 respecively and subtracting the component of their contribution from the measurement value. The HS-100 samples multiple data points around T0 and T2. The slopes are calculated with LSM (Least Square Method) by hardware DSP technique.

Another way of reducing the error from the leakage current is to shorten time interval between T0 and T2. With HS-100, T0-T2 time interval can be as short as 5-10usec. This means 0.5-1fF measurement error under a leakage current of 100pA and Vstep=1V even without leak cancelation. Interconnect leakage current is usually much smaller than 100pA. We can get enough accuracy without the leak cancellation technique described above.

## PARALLEL MEASUREMENTS RESULT

In this section, we will show experimental result of measuring 8 capacitors in parallel. Each device consists of 2 layers of electrodes as a capacitor. The device we measure does not have guard ring. This device only concerns about environmental changes of the capacitance. Hence, no-guard-ring design is not a problem for non-parallel measurement, but it is for parallel measurements. We need to avoid coupling between adjacent devices with special measurement sequence shown in Figure 4. Each capacitor is pulsed and measured in different moment of time. Because one measurement cycle is very fast, 8usec in this case, this does not cause severe throughput degradation. One set of measurements for 8 devices completes in 84usec. This sequence is repeated for averaging. This time, the device we measure does not have guard ring and there is rather large environmental noise. To get good noise rejection, measurements are repeated for 3 Power-Line-Cycles or 60msec. If the environment is clean and proper signal guarding is done, the measurement can complete much faster.

Table 1 shows measurement result. Absolute capacitance value of the each device is around few hundred Femto-Farads. In spite of the bad noise condition, measurement repeatability, or standard deviation, is around 1fF. Notable thing here is that difference between non-parallel and parallel measurements are less than 2.1fF for all devices. The pad layout of this experiment is depicted in Figure 5. Each DUT and its pads are adjacent to other DUTs and their pads. We can expect better parallel measurements result with guard ring.

This shows feasibility of on-chip parallel capacitance measurements using the DCM method. The HS-100 has up to 96 charge measurement channels. We can pursue further speed up with massively parallel measurements.

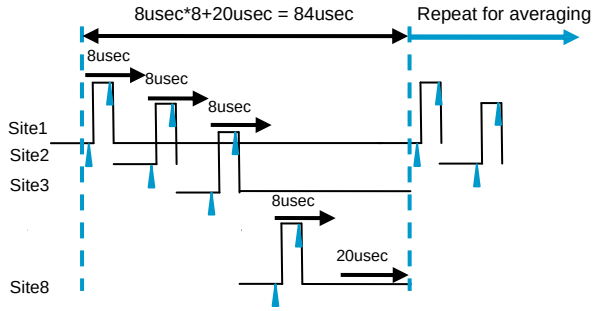



Figure4.Parallel Capacitance Measurements Sequence

TABLE 1
Parallel measurements deviation result

| | standard deviation of 20 measurements [F] | delta from non parallel measurement result [F] |
|---|---|---|
| site1 | 9.48E-16 | -5.92E-16 |
| site2 | 7.91E-16 | -4.79E-17 |
| site3 | 7.12E-16 | 2.07E-15 |
| site4 | 8.23E-16 | 1.65E-15 |
| site5 | 6.64E-16 | 1.25E-15 |
| site6 | 7.37E-16 | -1.78E-15 |
| site7 | 1.01E-15 | 9.23E-16 |
| site8 | 8.36E-16 | 2.07E-16 |

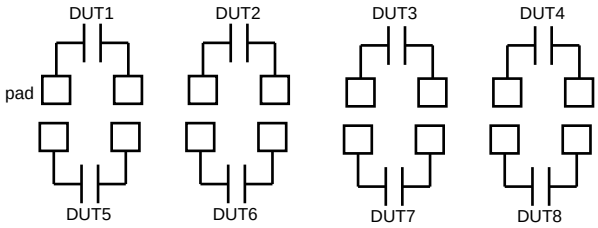



Figure5. Pad layout of 8 capacitors parallel measurements

## TFT ACTIVE MATRIX ARRAY MEASUREMENT

We also show TFT active matrix array measurement example. The target TFT devices are on a glass substrate. However, we think technique and result presented here can give some interesting insights to silicon test structure design.

Figure 6 shows schematic of TFT active matrix array. Each cell consists of a capacitor and a switching transistor. Although there is on-chip active device, the cell structure is simpler than that of CBCM. For TFT array testing, first, we charge each cell capacitor to a writing voltage, for example 1V, by scanning the matrix array switches (Figure 7). Then read the charge back at 0V by another scan (Figure 8). As we can see from Figure 8, voltage of a sense line doesn't change while reading the charge out. Therefore, only parasitic capacitance between the sense-gate line and gate capacitance of the switching transistor can cause measurement error. This means that coupling from non selected cells don't cause measurement error, thus, an accurate measurement is possible with a high density active matrix array. Moreover, the error from the gate-sense coupling is cancelled out by pulsing the gate voltage on and off, because the charge injected at off-on transition is discharged at on-off transition.

The measurement for each cell only takes 10-20usec. A TFT matrix array with 100,000 cells can be measured within a second by parallel measurements. Figure 9 shows spacial distribution of measured capacitances in gray scale and figure 10 shows its histogram. The cell capacitance varies between

1.22pF to 1.34pF and they have interesting spacial gradient on the array area of 25mm*35mm. If we design same kind of matrix array as a silicon test structure, spacial distribution of interconnect capacitances can be easily evaluated. This type of measurements and analysis has been done for transistor Vth or interconnect resistivity [10] [11] [12] [13] [14] [15] [16]. However, it has been difficult to get similar data for silicon interconnect capacitances.

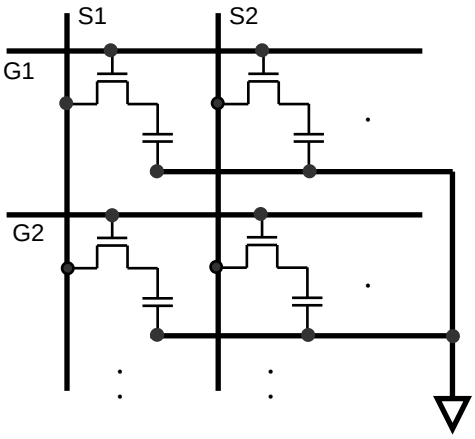


Figure6. Schematics of TFT active matrix array

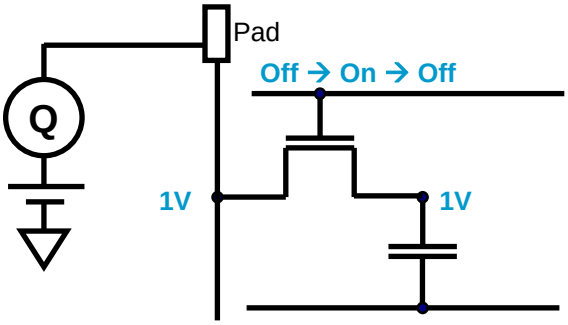


Figure7. Voltage condition while writing voltage

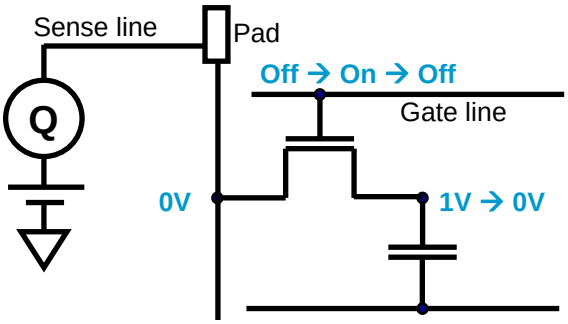


Figure8. Voltage condition while reading charge

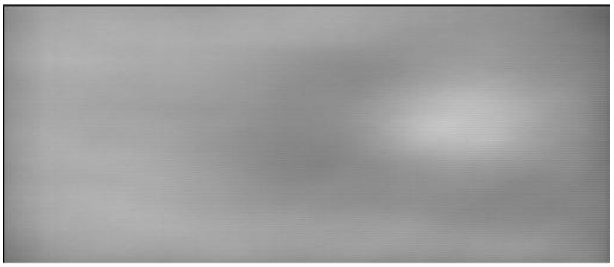


Figure9. TFT active matrix array capacitance spacial distribution on the array area of 25mm*30mm (capacitance value is from 1.22pF (black) to 1.34pF (white).)

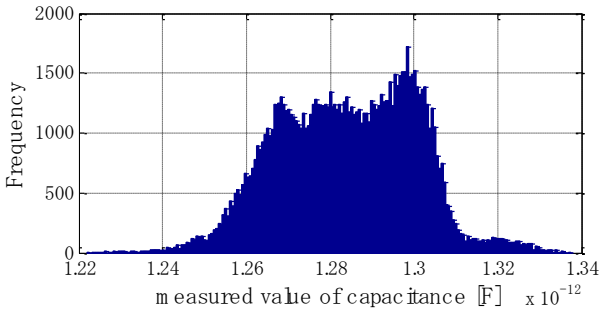


Figure10. Histogram of TFT active matrix array capacitance

## CONCLUSION

Direct charge measurement (DCM) is presented an interconnect capacitance verification method. Unlike CBCM, on-chip active device is not an essential necessity. It has Femto-Farad sensitivity and parallel measurements are achieved without adding significant error. With its' high resolution and throughput, drastic improvement of interconnect capacitance measurement efficiency is possible.

## REFERENCES

[1] Tomas Sutory, "METHODS FOR INTEGRATED CAPACITORS CHARACTERIZATION"

[2] James C. Chen, Bruce W.McGaughy, Dennis Sylvester and Chenming Hu, "An On-Chip, Attofarad Interconnect Charge-Based Capacitance Measurement (CBCM) Technique", IEDM 96-69

[3] Tetsuhisa Mido, Hiroshi Ito and Kunihiro Asada, "A Simple and Efficient Measurement Method for Characterizing Capacitance Matrix of Multilayer Interconnection in VLSI", IEEE transactions on Semiconducor Manufacturing, Vol 13, No 2 May 2000

[4] Ben Sell, Alejandro Avellan and Wolfgang Krautschneider, "Charge-Based Capacitance Measurements (CBCM) on MOS Devices", IEEE Transactions on Device and Materials Reliability, 2002,

[5] Yao-Wen Chang, et al, "Charge-Based Capacitance Measurement for Bias-Dependent Capacitance", IEEE Electron Device Letters, Vol 27 No 5 May 2006

[6] H.Brut, S.Martin and B.Froment, "Accuracy Improvement of the "Single Pattern Driver" Method

for the Characterization of Interconnect Capacitance in the Context of Nanometer Technology Development", IEEE ICMTS p303, March 2004
[7] Hui Sibg, Edwin Donis, Xi Qing Sun and K. R. Farmer, "Leakage Compensated Charge Method for Determining Static C-V Characteristics of Ultra-thin MOS Capacitors", Proceedings of the NIST International Conference on Characterization and Metrology for ULSI Technology, p231-234, 1998
[8] Zhenqui Nihg, Luc De Schepper, Henri-Xavier Delecourt, Renaud Gillon and Marnix Tack, "A Simple and Accurate Capacitance Ratio Measurement Technique for Integrated Circuit Capacitor Arrays", IEEE ICMTS Vol .18 p159 April 2005
[9] Jurgen Deveugele, Libin Yao, Michiel Steyaert and Willy Sansen, "Mismatch characterization of chip interconnect resistance", IEEE ICMTS, Vol18 p183 April 2005
[10] Frederic Boeuf, Manuel Sellier, Alexis Farcy and Thomas Skotnicki, "Impact of Layout, Interconnects and Variability on CMOS Technology Roadmap", 2007 Symposium on VLSI Technology Digest of Technical Papers, 3A-2
[11] Jerry Hayes, Kanak Agarwal and Sani Nassif, "Rapid Characterization of Parametric Distribution Using a Multi-meter", IEEE ICMTS p17 March 2008
[12] Michihiro Kanno, et al, "Empirical Characteristics and Extraction of Overall Variations for 65-nm MOSFETs and Beyond", 2007 Symposium on VLSI Technology Digest of Technical Papers, 6A-2
[13] Christopher Hess, et al, "Device Array Scribe Characterization Vehicle Test Chip for Ultra Fast Product Wafer Variability Monitoring", IEEE ICMTS, p145, March 2008
[14] Kanak Agarwal, Sani Nassif, Frank Liu, Jerry Hayes and Kevin Nowka, "Rapid Characterization of Threshold Voltage Fluctuation in MOS Devices", IEEE ICMTS p74, March 2007
[15] Ulrich Schaper, Jan Einfeld and Anke Sauerbrey, "Parameter Variation on Chip-Level", IEEE ICMTS Vol.18 p155, April 2005
[16] Masaharu Yamamoto, Yayoi Hayasi, Hitoshi Endo and Hiroo Masuda, "Measurement of inner-chip Variation and Signal Integrity by a 90-nm Large-scale TEG", IEEE ICMTS p207, April 2005
[17] Muthu Karthikeyan, et al, "Short-flow Test Chip Utilizing Fast Testing for Defect Density Monitoring in 45nm", IEEE ICMTS p56, March 2008
[18] Muthu Karthikeyan, et al, "A 65nm Random and Systematic Yield Ramp Infrastructure Utilizing a Specialized Addressable Array with Integrated Analysis Software", IEEE ICMTS p106, 2006
[19] Christopher Hess, et al, "Scribe Characterization Vehicle Test Chip for Ultra Fast Product Wafer Yield Monitoring", IEEE ICMTS p110, 2006
[20] Yasuhiro Miyake, Kiyoshi Chikamatsu, Masaharu Goto, Junichi Mizoguchi, "The New Electrical Test Method for LCD Cell Manufacturing Process", Journal of SID, Vol.16, p83, 2008
[21] Yasuhiro Miyake, Kiyoshi Chikamatsu, Junichi Mizoguchi, "Intoroduction of Defect Analysis by Cell Electrical Test for LTPS LCD and OLED", IDW'07 Digest, p157, 2007
[22] Robert Wisnieff, et al, SID90 Digest, p190, 1990
[23] Yasuhiro Miyake, Junichi Mizoguchi, "New Electrical Test Method for on-wafer MEMS Devices", Proceedings of the IEEJ 25th Sensor Symposium, p274, 2008

# 4K-cells Resistive and Charge-Base-Capacitive Measurement Test Structure Array (R-CBCM-TSA) for CMOS Logic Process Development, Monitor and Model

Kelvin Y.Y. Doong[a], Keh-Jeng Chang[b], S.-C. Lin[a], H. C. Tseng[a], Akis Dagonis[a], Samuel Pan[a]
No. 8, li-Hsin Sixth Rd., Hsin-Chu, Taiwan, R.O.C., Tel. +886-3-5636688 ext. 6855,e-mail: kelvin_doong@tsmc.com
[a]Taiwan Semiconductor Manufacturing Company, Science-Based Industrial Park, Taiwan
[b]Department of Computer Science, National Tsing Hua University, Hsin-Chu, Taiwan

## Abstract

To maximize the design efficiency of the test chip area and maintain the high accuracy measurement requirement of resistors and capacitors, a 4K-cells resistive and charge-base capacitive test structure array is designed for CMOS logic process development, monitor and model. The test chip utilizes 4-terminal (one of 4 is strongly grounded) Kelvin force/sense measurement for resistive-type and charge-base capacitance measurement (CBCM) for capacitive-type test structures. With the aid of memory-addressing design scheme, any one of the device-under-test in an array can be randomly or sequentially selected for testing with all of them sharing a common probe pad group. To accelerate the testing speed, the address control signals of 8 test structure array are connected in parallel for synchronized parallel testing. A 32x16x8 test structure array has been implemented by utilizing a state-of-the-art logic process to demonstrate design feasibility. The results confirm the excellence of this architecture in measurement with 0.1fF for capacitive and 0.1ohm for resistive systematic errors, and 7 times testing speed improvement.

## I. Introduction

The array-type test chip has been widely adopted because of its high area utilization rate since Buehler M. G. *et al.* introduced the array-type test chip in 1981 [1]. To resolve and minimize the parametric variation induced by process variation as well as spatial effect and enable the defect detection capability of high spatial resolution, the array-type design was used to characterize the electrical performance of the various types of test structure such as transistors, diodes, resistors, capacitor [2]-[14]. Only few of them can be used for capacitor measurements [12], [13]. To address the "soft failing", Hess et al. employed a high density resistive-type array with 4-termainl Kelvin force/sense [14]. The aforementioned test chip either focused on high density array of transistors or resistors. None of them provides an array-type test chip with high accuracy measurement and high array packing density for both of resistive-type and capacitive-type test structures.

This work provides a feasible solution, referred as resistive and charge-base capacitive measurement test structure array (R-CBCM-TSA) for characterizing process variation with the following features:

- A higher test structure density compared with conventional pad-addressing/connecting test structures.
- High parametric accuracy.
- High testing throughput with parallel testing capability (compared with conventional parametric tester, there are at least two times throughput grain).

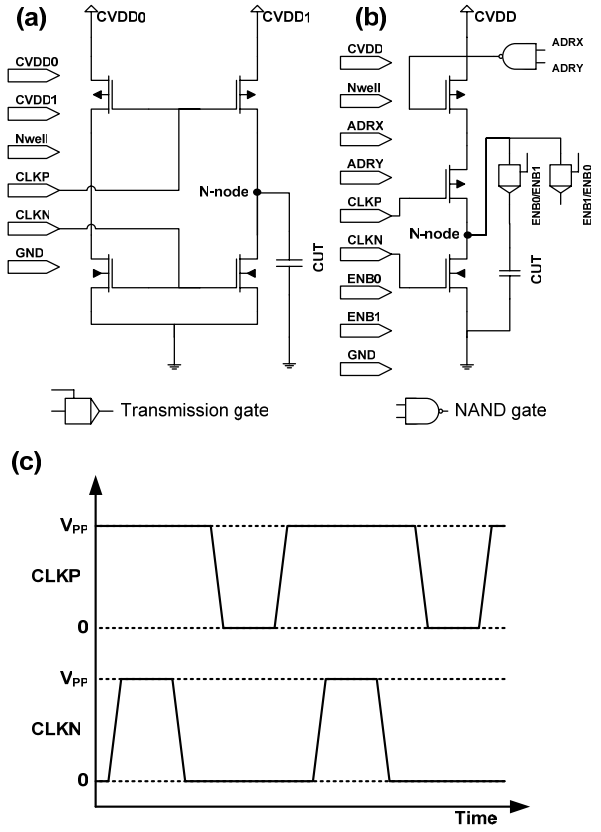



Fig. 1: Schematic of charge-base capacitance measurement (CBCM) cell. (A) Conventional cell. (b) Array cell. (c) Clock waveform.

- Testable with in-line parametric test hardware
- Spatial localization of failure events.
- Configurable test cell-size.
- Extensible to implementation as a production monitor.

Section-II describes the test structure array design methodology. The proposed design is verified and characterized. In Section-III, process parametric variation of metal-line and overlay-shift are discussed and a summary is presented in Section-IV.

## II. Cell/Array Design and Verification

To accommodate the resistive-type and capacitive-type test structure into an array cell, the conventional charge-base capacitance measurement test structure was modified from two pseudo inverters into a circuit composed of a PMOS controlled pseudo inverter and two transmission gates for selecting either **C**apacitor **U**nder **T**est (CUT) or **R**eference **C**apacitor **U**nder **T**est (ReCUT) as depicted in

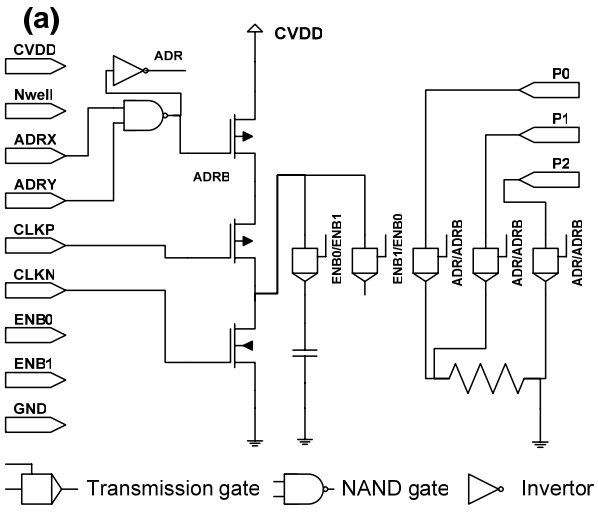



Fig. 2: (a) Schematic of R-CBCM cell. (b) Physical layout of R-CBCM cell.

Fig. 1. The $C_{ReCUT}$ is composed of $C_{GD,NMOS}$, $C_{GS,PMOS}$, $C_{Juncation,PMOS}$, $C_{Junction,NMOS}$ and interconnect parasitic capacitance including contact-to-gate capacitance ($C_{CO}$), where $C_{GD}$ and $C_{GS}$ mean the capacitance of gate-to-drain and gate-to-source, respectively. For the new proposed CBCM array cell, the $C_{ReCUT}$ will include extra capacitance resulted from $C_{GD}$, $C_{GS}$ and $C_{CO}$ of two transmission gates.

Compared with the conventional and new proposed CBCM cells, there are 6 and 9 probe pads for testing where the new proposed can commonly share 9 out of 9 probe pads but the conventional one can only share 4 out of 6. Besides of area saving through probe pads sharing, the new proposed one can eliminate the charge-injection error induced by the pseudo inverter switching. The charge-injection error is mainly contributed from charge redistribution during pseudo inverter's PMOS/NMOS is switched on/off, where the charge at the channel region will be drained out or sunk into N-node. Because of only a pseudo inverter used in the new proposed CBCM, the charge-injection errors can be eliminated through two-pass measurements.

When ENB0 is set to on and ENB1 is set to off, the charge current through PMOS is expressed as

$$I_0=(C_{CUT}+C_{ReCUT})*V_{PP}*f \quad (1)$$

Where $V_{PP}$ is the voltage of voltage source connected to

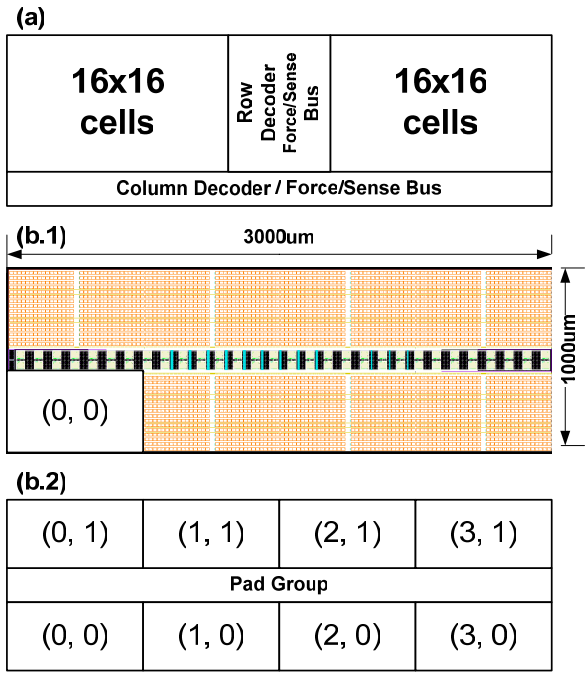



Fig. 3: RC array test chip. (a) 32x16 sub-chip diagram. (b.1) bird-view. (b.2) Block diagram.

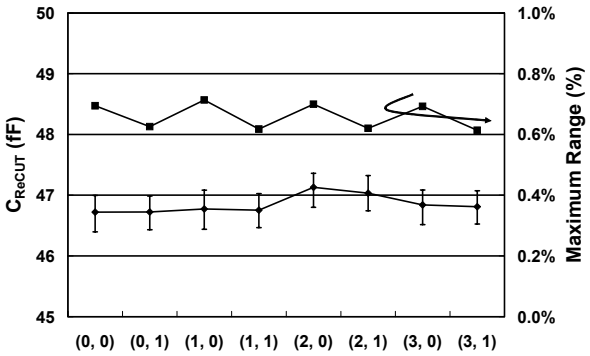



Fig. 4: $C_{ReCUT}$ plotted as a function of test chips located as the various locations.

CBCM and $f$ is the frequency of non-overlap clock. When ENB0 is set to off and ENB1 is set to on, the charge current through PMOS is expressed as

$$I_1=C_{ReCUT}*V_{PP}*f \quad (2)$$

$$C_{ReCUT}=I_1/V_{PP}/f \quad (2.1)$$

Using the two-pass measurement, the $C_{CUT}$ can be extracted by substituting Eq. (2) into Eq.(1) as expressed in Eq. (3)

$$C_{CUT}=(I_0-I_1)/V_{PP}/f \quad (3)$$

To measure the resistor and capacitor within the same test structure cell, the transmission gates are designed into CBCM cell. To minimize the probe pad usage, a modified 4-port Kelvin force/sense test structure was enacted in this design. The modified 4-port Kelvin test structure is as same as the conventional one except one of 4-port was strongly grounded. For the detail design optimization of resistive test structure array, it could be found in [11]. The schematic of R-CBCM cell was illustrated in Fig. 2, where the schematic is shown in (a) and the physical layout of cell size 25.88x22.34 $um^2$ is presented in (b).

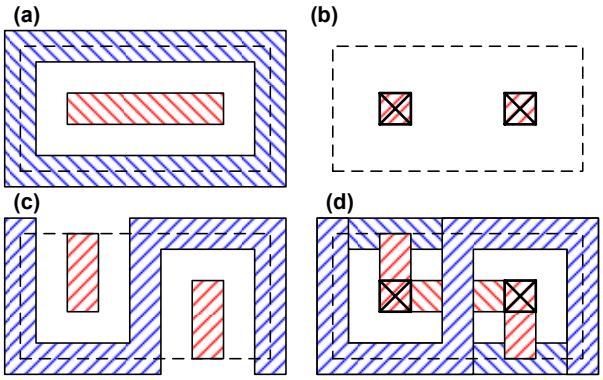



Fig. 5: Physical layout of overlay-shift resistive-detection test structure . (a) Under-layer. (b) Via-layer. (c) Top-layer. (d) Basic cell, where the blue represents the ground connection and the red is connected to $V_{DD}$.

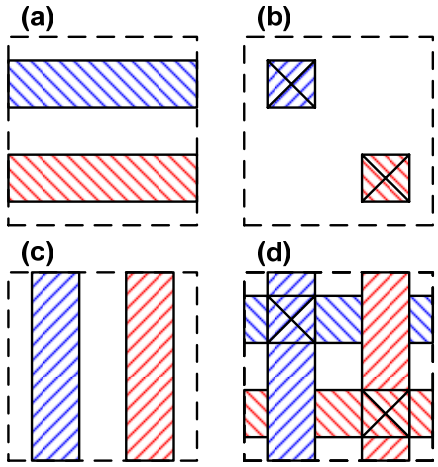



Fig. 6: Physical layout of overlay-shift capacitve-detection test structure. (a) Under-layer. (b) Via-layer. (c) Top-layer. (d) Basic cell, where the blue represents the ground connection and the red is connected to $V_{DD}$.

A 32x16 R-CBCM test structure array is presented in Fig. 3(a). The R-CBCM-TSA consists of M-bit and N-bit decoders at the X- and Y-direction, bit-cell array and in

put/output (I/O) bus for CBCM, current-force, and voltage -sense signals. M-bits and N-bits decoders used as X- and Y-decoders are implemented to select the targeted bit-cell. In order to enable the parallel testing capability, there are 8 R-CBCM-TSAs connected in parallel as depicted in (b), where the address signals are commonly connected and the CBCM/4-port-Kelvin's current/voltage force/sense ports are electrically isolated.

The results of 32x16 R-CBCM array cell measurement at the 8 various locations with a test chip are shown in Fig. 4, where the mean value of $C_{ReCUT}$ is around 46.7fF to 47.1fF and the maximum range is around 0.6~0.7% for all of 8 sub-dies. It indicated that R-CBCM has a high degree of capacitive measurement resolution down to 0.1fF.

There are two devices in an R-CBCM cell; one is a resistor and the other is a capacitor. For the resistor measurement, a two-pass (forward/reverse current flow) 4-Kelvin force/sense measurement with fixed current ($I_0$) was

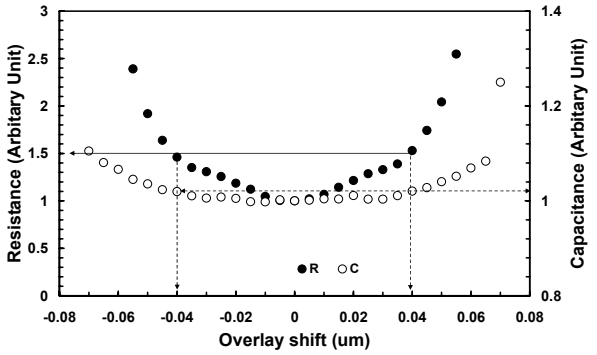



Fig. 7: $C_{CUT}$ plotted as a function of the overlay-shift at the X-direction.

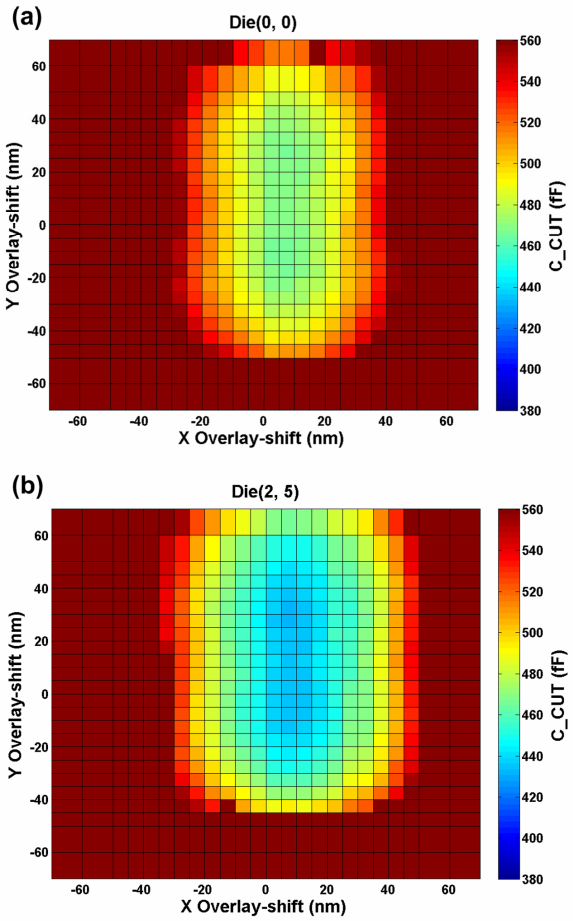



Fig. 8: $C_{CUT}$ contour plotted as a function of the various overlay-shifts.

performed to eliminate the thermal heating.

$$\Delta V_{forward}=R_{DUT}*I_{forward}+V_{thermal}(I_{forward}) \quad (4)$$

$$\Delta V_{reverse}=R_{DUT}*I_{reverse}+V_{thermal}(I_{reverse}) \quad (5)$$

, where $\Delta V_{forward}$ and $\Delta V_{reverse}$ are voltage drops at two sense ports for the forward and reverse current flow, respectively; $R_{DUT}$ is the DUT resistance; $I_{forward}$ is equal to $I_0$ and $I_{reverse}$ is equal to $-I_0$; $V_{thermal}(I_{forward})$ *and* $V_{thermal}(I_{reverse})$ are the voltage deviation induced by thermal heating during measurement. Since the voltage deviation is independent of current flow direction, $V_{thermal}(I_{forward})$ should be almost equal to $V_{thermal}(I_{reverse})$. Thus, by combining Eq.(4) and Eq.(5), the

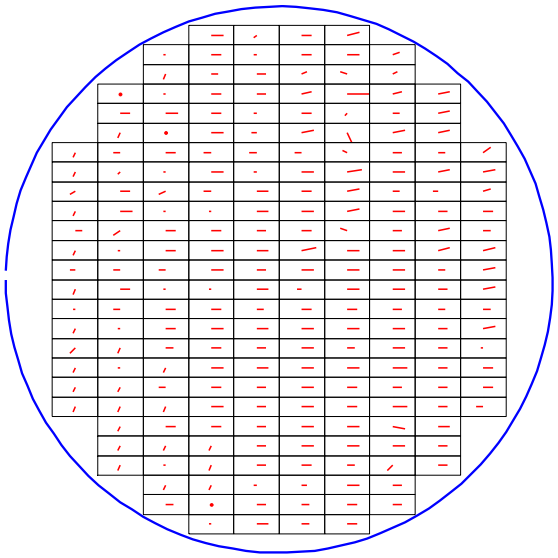


Fig. 10: Wafer map of overlay-shift, where each of rectangle represents -20nm~20nm overlay-shift at the X-direction and Y-direction.

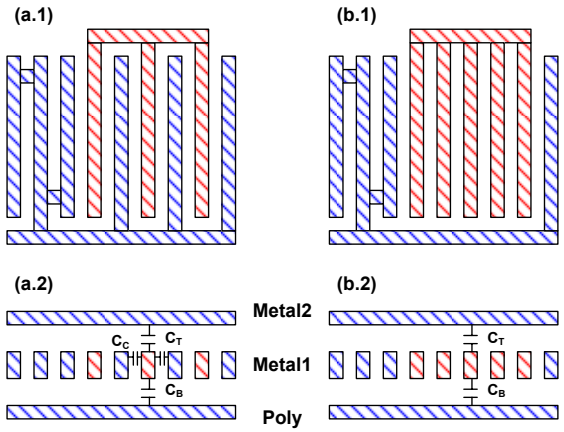



Fig. 11: Physical layout plain view of (a.1) intra-layer (b.1) inter-layer capacitance measurement test structure, where there is a 4-port Kelvin force/sense test structure located at the right-hand side. Cross section view of (a.2) intra-layer (b.2) inter-layer.

DUT resistance can be expressed as follows:

$$R_{DUT}=(\Delta V_{forward}-\Delta V_{reverse})/2I_0 \quad (6)$$

As the design requirement is specified in Sec. I, R-CBCM-TSA is designed with 8 parallel testing capability. To test all of R-CBCM in 8 test structure arrays (there are 4096 4-port Kelvin and 4096 CBCM testing), it takes 14 minutes for both of in-line parametric testers such as Agilent-407x series or Keithley-S680. Besides of using conventional parametric tester, an instrument-based in-house tester was employed to test this test chip and it takes only 2 minutes, which achieves 7 times throughput improvement.

## III. Experimental Results

The following describes two examples in process control, monitor and model, one is the electrical overlay-shift of Via-1 chains, and the other is interconnect (Metal-2) capacitance variation characterization. The design infrastructure

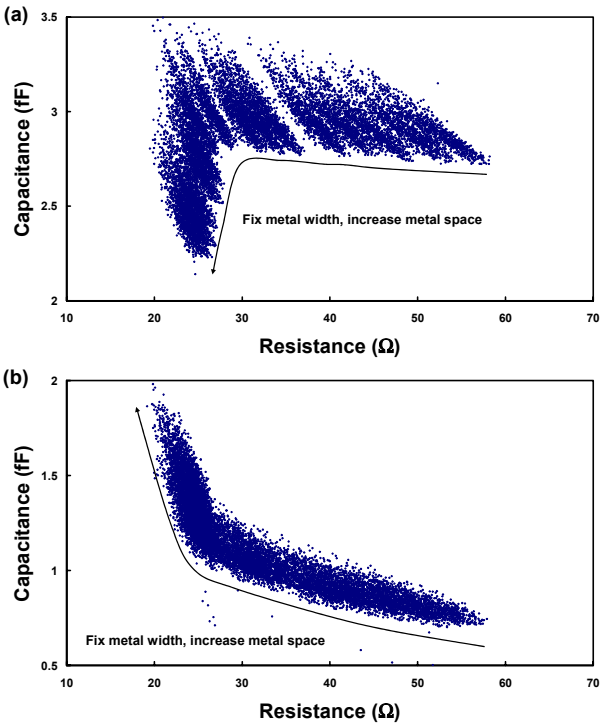



Fig. 9: Interconnect capacitance plotted as a function of resistance through the various pitches, where the metal width is fixed at the nominal and the space is varied from the nominal to the wide space. (a) the intra- and inter-layer capacitance (b) the inter-layer capacitance.

described in [14] was used to generate hundreds of test structure sets.

The Via1 chains test structure is shown in Fig. 5 and the comb with Via1 array is in Fig. 6. The metal strip of Metal-1 and Metal-2 are orthogonally placed and Via1 is diagonally located at the cross area. For Via1 chains test structure, the metal strip is surrounded with the uniform runner. To detect the electrical overlay shift, the Via1 location is intentionally drawn in the way of relative shifted to the under layer metal. A complete set of electrical overlay-shift test structure contains 625 (25x25) permutation structures where the overlay conditions are varied from -70nm to 70nm both at the X- and Y-direction. With the feature of this test structure, the Via1-to-Metal or the Metal2-toVia1 overlay shift will increase the total capacitance, thus, the overlay shift signal can be translated into the capacitance signal.

The e-test of the overlay-shift test structures of Fig. 5 is shown in Fig. 7. It indicates that the sensitivity of Via1 chains resistance to overlay-shift is much more than the one of Via1 chains capacitance to overlay-shift. The e-test of the overlay-shift test structures of Fig. 6 is shown in Fig. 8, where (a) presents well aligned structure and (b) illustrates abnormally aligned structures. With the aid of high spatial resolution (5nm) in the test structure design, the e-test data reveals that there is a 30nm overlay-shift process window available. The wafer map of systemic overlay-shift is profiled in Fig. 10, where the systemic overlay-shift is extracted from the lowest capacitance within 625 test structures and the range at the X- and Y-direction is ±20nm.

The interconnect(Metal-2) capacitance variation characterization test structure is shown in Fig. 11, where (a)

is used to extracted the intra- and inter-layer capacitance, and (b) is for the inter-layer capacitance. Accordingly, the total capacitance of (a) and (b) can be expressed as follows:

$$C_{Total,a}=(2n*C_C+n*C_T+n*C_B)*L+\alpha \quad (7)$$

$$C_{Total,b}=(2*C_C+2*n*C_T+2*n*C_B)*L+\beta \quad (8)$$

, where $C_C$ represent the intra-layer capacitance in the unit of fF/um; $C_T$ and $C_B$ are the inter-layer capacitance relative to the top-layer and the bottom-layer in the unit of fF/um; $\alpha$ and $\beta$ the parasitic capacitance of test structure; n is the teeth number of comb test structure in Fig. 11 (a) or (b). Both of $C_{Total,a}$ and $C_{Total,b}$ are measured by the CBCM and corresponding resistance is measured at the same test structure. The $C_{Total,a}$ and $C_{Total,b}$ versus corresponding resistance are shown in Fig. 9(a) and (b), respectively, where the intra-layer capacitance ($C_C$) and the inter-layer capacitance ($C_T$ and $C_B$) can be extracted from Eq.(7) and Eq.(8)

## IV. Conclusion

A new test structure array design called resistive and charge-base capacitive measurement test structure array (R-CBCM-TSA) is implemented with 8 parallel testing channel. An optimal design methodology accounting for high resolution of resistance and capacitance measurement has been discussed and measurement with conventional parametric tester and ad hoc parallel tester has been demonstrated, where 7 times throughput gain compared with the conventional approach.

By utilizing the R-CBCM-TSA, area design efficiency is improved by a factor of 100~250 over conventional test structures. The capabilities reported in this work include 5nm spatial resolution which can be electrically characterized for overlay process control. In addition, local fluctuation in interconnect can be statistically profiled and modeled. The variation source decomposition also confirms that the proposed R-CBCM-TSA design is an essential tool for process monitor and control, and the variability diagnosis for the further development of sub 45nm node technology.

## V. References


[1] Buehler M. G. et al., "Role of Test Chips in Coordinating Logic and Circuit Design and Layout Aids for VLSI," *Solid State Technology*, pp. 68-74., Dec. 1981

[2] T. Tsunomura, A. Nishida, F. Yano, A.T. Putra, K. Takeuchi, S. Inaba, S. Kamohara, K. Terada, T. Hiramoto, T. Mogami, "Analyses of 5σ Vth Fluctuation in 65nm-MOSFETs Using Takeuchi Plot," in *dig. Tech. Papers 2008 Symp. VLSI Circuit*, June 16-18, 2008

[3] K. Agarwal, F. Liu, C. McDowell, S. Nassif, K. Nowka, M. Palmer, D. Acharyya, and J. Plusquellic, "A test structure for characterizing local device mismatches" *Proc. Symp. VLSI Circ.*, June 2006.

[4] B. Smith, U. Annamalai, A. Arriordaz, V. Kolagunta, J. Schmidt, M. Shroff, "A Novel Biasing Technique for Addressable Parametric Arrays", *IEEE International Conference on Microelectronics Test Structures*, pp. 166-171, March 2008.

[5] R. Kuroda, A. Teramoto, T. Komuro, W. Cheng, S. Watabe, C.F. Tye, S. Sugawa, T. Ohmi, "Characterization of MOSFETs Intrinsic Performance using the In-Wafer Advanced Kelvin-Contact Device Structure for High Performance CMOS LSIs", *IEEE International Conference on Microelectronics Test Structures*, pp. 155-159, March 2008.

[6] C. Hess, M. Squcciarini, S. Yu, J. Burrows, J. Cheng, R. Lindley, A. Swimmer, S. Winters, "High Density Test Structure Array for Accurate Detection and Localization of", *IEEE International Conference on Microelectronics Test Structures*, pp. 131-136, March 2008.

[7] F. Rigaud, J.M. Portal, H. Aziza, D. Nee, J. Vast, F. Argoud, B. Borot, "Mixed Test Structure for Process Variation and Hard Defect Detection", *IEEE International Conference on Microelectronics Test Structures*, pp. 52-55, March 2008.

[8] J. Hayes, K. Agarwal, S. Nassif, "Rapid Characterization of Parametric Distributions Using a Multi-meter", *IEEE International Conference on Microelectronics Test Structures*, pp. 17-20, March 2008.

[9] A. Mizumura, T. Suzuki, T. Arima, H. Maeda, H. Ammo; "A Study of Variation in Characteristics and Subthreshold Humps for 65-nm", *IEEE International Conference on Microelectronics Test Structures*, pp. 8-10, March 2008.

[10] B.L. Ji, D.J. Pearson, I. Lauer, F. Stellari, D.J. Frank, L. Chang, M.B. Ketchen, "Operational Amplifier Based Test Structure For Transistor Threshold Voltage", *IEEE International Conference on Microelectronics Test Structures*, pp. 3-7, March 2008.

[11] Kelvin Y.Y. Doong, Jim Bordelon, L.J. Hung, C.C. Liao, S.C. Lin, P.S. Ho, Sunnys Hsieh, and K.L.Young, "Field-Configurable Test Structure Array (FC-TSA): Enabling Design for Monitor, Model, and Manufacturability", *IEEE trans. on semiconductor manufacturing*, Vol. 21, No. 2, pp. 169 – 179, May, 2008

[12] Shin-ichi Ohkawa, Masakazu Aoki, Hiroo Masuda, "Analysis and Characterization of Device Variations in an LSI Chip Using an Integrated Device Matrix Array", *IEEE Trans. Semiconductor Manufacturing*, Vol. 17, no. 3, pp. 248 – 254, Aug., 2004.

[13] *R. Lefferts*, *C. Jakubiec*; *"An Integrated Test Chip for the Complete Characterization and Monitoring of a 0.25gm CMOS Technology that fits into Five Scribe Line Structures"*, IEEE International Conference on Microelectronic Test Structures, pp. 59-63, March 2003

[14] C. Hess, M. Squcciarini, S. Yu, J. Burrows, J. Cheng, R. Lindley, A. Swimmer, S. Winters; "High Density Test Structure Array for Accurate Detection and Localization of", *IEEE International Conference on Microelectronics Test Structures*, pp. 131-136, March 2008.

[15] Kelvin Y.Y. Doong, Jurcy C.H. Huang, S.C. Lin, L.J. Hung, Sunnys Hsieh, Robin C. J. Wang, P.S. Ho, Philips Lin, Roger Kang, Chia-Chi Chu, and K.L.Young, "Infrastructure Development and Integration of Electrical-based Dimensional Process-Window Checking", *IEEE trans. on* semiconductor manufacturing, Vol. 17, No. 2, pp. 123 – 141, May, 2004.

# Practical Considerations for Measurements of Test Structures for Dielectric Characterization

Wenbin Chen, Kevin G. McCarthy and Alan Mathewson

Tyndall National Institute and Dept. of Electrical & Electronic Engineering,
University College Cork, Ireland

## ABSTRACT

This paper presents a method for measuring the complex permittivity of dielectric material on a dielectric/metal stack. A series of circular capacitor and transmission line test structures are designed and fabricated. The methodology has been verified by measuring the dielectric constant of a known $SiO_2$ layer using Capacitance-Voltage (C-V) measurement and scattering parameter (S-parameter) measurements. The combination of C-V measurement and S-parameter measurement is shown to be suitable for characterization of dielectric material on the complex cross-sections.

## INTRODUCTION

High dielectric constant material thin films are now commonly used in microelectronic devices. The characteristics, such as dielectric constant and dielectric loss tangent, directly determine the quality of the devices and systems. Before high-k materials can be used in conjunction with modern Si processes for circuit design, the materials must be accurately characterized.

An advanced test structure design methodology for dielectric characterization of novel high-k materials has been presented previously [1] [6]. In order to extract the dielectric constant ($\varepsilon_r$), a number of different coplanar-waveguide structures and circular capacitor test structures have been designed. For initial characterization, the circular capacitor test structure has been found to be the most straight-forward to interpret. The main difficulty in verifying a particular wafer probe characterization method is the lack of an available standard which can be used for reference purposes.

## TEST STRUCTURE

Test structures on a known multilayered silicon substrate (Fig. 1) can be used as a standard to test the characterization technique and also to test the accuracy of the measurement. The main test structure is based on a circular capacitor which requires only the top metal layer to be patterned [2]. In the cross-section shown in Fig. 1, the top $SiO_2$ layer (thickness $t_D$) would normally be replaced by the high-k dielectric to be characterized. Here, $SiO_2$, a known dielectric, is used to verify the measurement techniques. As seen, this $SiO_2$ layer has a blanket layer of Pt underneath and a pattered layer of Pt (and Au, not shown) on top.

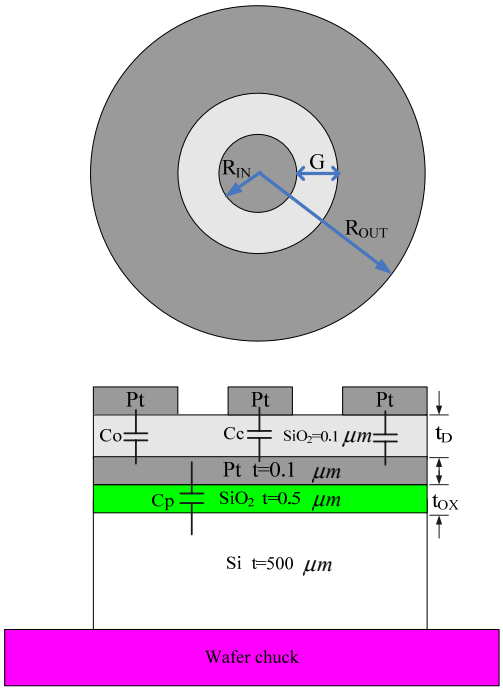



Fig. 1. A top view and a cross section of the circular capacitor test structure.

In Fig. 1, one probe makes contact to the centre circular metal disc ($R_{IN}$ = 65µm, G = 20µm) which functions as the top electrode of the capacitor, and the second probe contacts the metal outer ring ($R_{OUT}$ = 600µm) which functions effectively as the ground of the capacitor. The equivalent electrical circuit of the measurement is illustrated in Fig. 2. This is based on the circuit presented in [3].

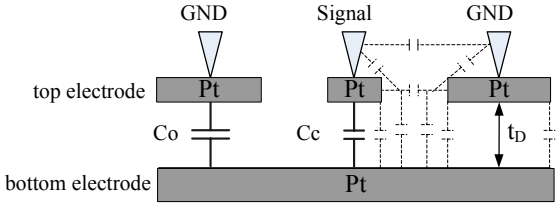



Fig. 2. A cross section of the test structure with 3 probes. The parasitic capacitances are indicated with striped grey colors [3].

978-1-4244-4259-1/09/$25.00 ©2009 IEEE

The capacitance between the centre disc and bottom electrode ($C_C$) is in series with the capacitance between the outer ring and bottom electrode ($C_O$). Since $C_C$ is much smaller than $C_O$, the response of the whole circuit (Fig. 3) is dominated by the smaller capacitor [2].

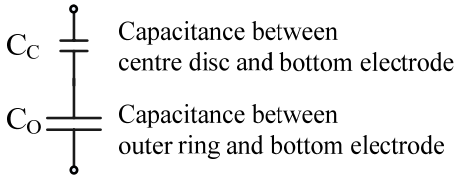



Fig. 3. Ideal equivalent circuit of the test structure.

$$\frac{1}{C} = \frac{1}{C_C} + \frac{1}{C_O} \qquad (1)$$

Therefore the measured capacitance, C, is equal to $C_C$, if $C_O \gg C_C$.

The distance between the contact paths is in the µm-range, while the dielectric thickness is in the nm-range ($t_D$ = 100nm), so the electric field is screened by the bottom electrode [3]. The capacitance from the ground path through the dielectric is neglected. The capacitances between the ground and signal probes are negligible because they have been shown to be in the low fF-range by simulation [3]. Thus, the influence of all parasitic capacitances is negligible and these are ignored [3].

## C-V Extraction Techniques

Generally, C-V measurements are straight-forward and allow fast extraction of $\varepsilon_r$. A Hewlett Packard 4280A C-V Plotter integrated with a SUSS MicroTec probe station is used to carry out the measurements.

The capacitance $C_C$ is calculated as follows:

$$\begin{aligned} C_C &= \frac{\varepsilon_O \varepsilon_r}{t_D} \cdot A_C \\ &= \frac{\varepsilon_O \varepsilon_r}{t_D} \cdot \pi \cdot (R_{IN})^2 \\ &= C_{DO} \cdot \pi \cdot (R_{IN})^2 \end{aligned} \qquad (2)$$

Where $C_{DO} = \varepsilon_O \varepsilon_r / t_D$ is the capacitance per unit area, $\varepsilon_o$ is the permittivity of free-space, $\varepsilon_r$ is the relative permittivity (dielectric constant) of the dielectric layer, $t_D$ is the thickness of the dielectric layer and $A_C$ is the area of the centre disc.

The capacitance was measured for each sample with the probes lifted in the air and also with the probes touching the metal patterns [4]. The capacitance measured with the probe in the air is then subtracted from the measurements when the probes are contacted to the sample.

When the capacitance for a certain structure is measured, the dielectric constant of the material can be found by rearranging the capacitance formula i.e.

$$C = \frac{\varepsilon_O \varepsilon_r}{t_D} \cdot A_C \Rightarrow \varepsilon_r = \frac{C}{A_C} \cdot \frac{t_D}{\varepsilon_O} \qquad (3)$$

Initial measurements with the high-k dielectric were different to the expected results and a method to verify the measurement setup was investigated. Because many patterns were available on the test structure, it was possible to measure the capacitance between two large discs (diameter 600µm) as shown in Fig. 4. The electrical equivalent of the cross-section is shown in Fig. 5. Initially, the equivalent circuit was assumed to be two equal value capacitors in series (Fig. 5 (a)). But measurements again produced a much smaller capacitance than expected. In order to fully understand the behavior, new test samples were fabricated with a known dielectric layer ($SiO_2$) and an equivalent sub-circuit model to describe the measurement was investigated. The new sub-circuit model is shown in Fig. 5 (b).

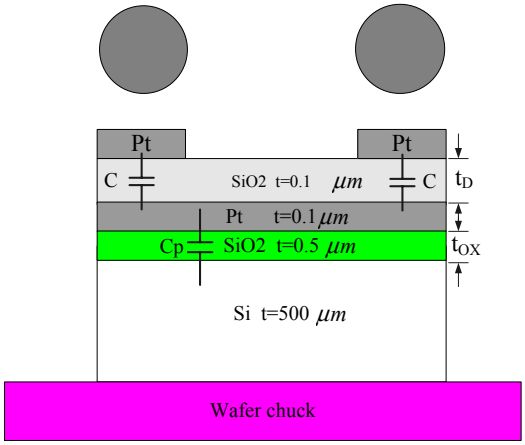



Fig. 4. Measurement setup for two series capacitors.

It was discovered during measurements that the setup had a large influence on the measured capacitance values. From Fig. 4 it is seen that there is an unwanted capacitance $C_p$ formed by the oxide layer between the bottom Pt layer and the silicon substrate. When the silicon substrate is mounted onto a wafer holder (chuck) for measurement, this will influence the measurement results.

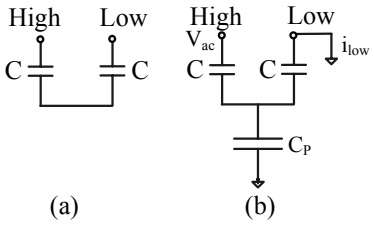



Fig. 5. Electrical equivalent of two series capacitors. (a) Ideal case (b) Real setup with chuck grounded.

To investigate if the parasitic capacitance $C_p$ was influencing the measurements, the capacitance between the two top discs was measured under different conditions as summarized in Table I.

TABLE I
CAPACITOR MEASUREMENTS UNDER DIFFERENT CONDITIONS

| Case | Chuck grounded | Chuck-floating | Chuck on insulator |
|---|---|---|---|
| $C_{meas}$ | 148.419pF | 167.402pF | 202.108pF |

For the case where the samples are mounted directly on the chuck and with the chuck grounded, it is seen that the measured capacitance is the smallest of all the configurations. This corresponds to $C_p$ having a large influence on the measurements. The influence of $C_p$ is reduced by first disconnecting the chuck and then moving the samples away from the conducting chuck by placing them on an insulator (a petri dish was used as the insulator). The latter measurement condition is referred to as the "insulator" method later in the paper.

As a result of the series of tests, it is considered that the most reliable capacitance measurements between two discs on the top surface of the samples are when the samples are mounted on an insulator with the chuck floating (Fig. 6).

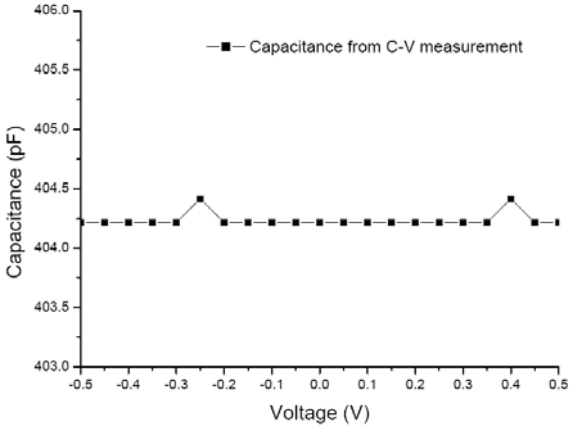



Fig. 6. Measured capacitance using 4 mm high insulator placed on chuck to physically separate the samples and the chuck.

The resulting value of $\varepsilon_r$ is shown in Fig. 7.

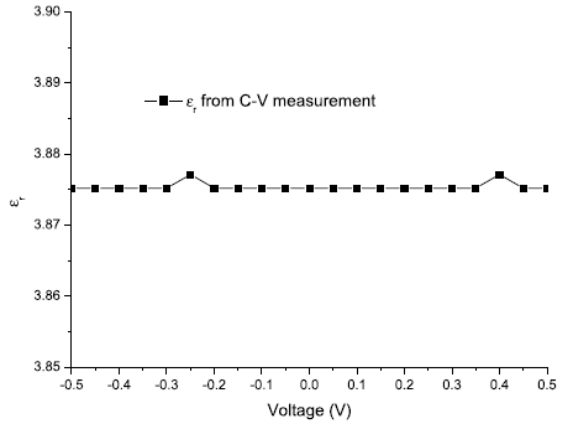



Fig. 7. Extracted $\varepsilon_r$ from C-V measurement of Fig. 6.

The insulator between the wafer and the chuck reduces the complicated coupling between the wafer chuck and the ground, and results in improved measurements at higher frequencies [5].

Next, an understanding of the behavior of the C-V measurements was sought by means of a circuit analysis of Fig. 5 (b). This shows the connections of the 'High' and 'Low' terminals of the C-V meter. Although, the C-V meter operates an elaborate current balancing principle in practice, its operation can be represented by the application of an ac excitation on the 'High' terminal and the measurement of the resultant ac current at the 'Low' terminal. Any stray capacitance to ground ($C_p$) effects $i_{Low}$ and thus the apparent measured capacitance. An electrical analysis of Fig. 5 (b) shows that the apparent capacitance between the 'High' and 'Low' terminals will be

$$C'_{meas} = \frac{C}{2 + \frac{C_p}{C}} \qquad (4)$$

TABLE II
ELEMENTS OF THE SUB-CIRCUIT

| C | $C_p$ | $C'_{meas}$ | $C_{meas}$ (Chuck grounded) |
|---|---|---|---|
| 404.216pF | 302.44pF | 147.084pF | 148.419pF |

Table II shows measurements and calculations to verify this model. Here, C is the capacitance of the discs to the underlying Pt layer determined from the "insulator" method. Once C is known, $C_p$ is determined by measuring the capacitance of a top disc to the grounded chuck (which is just C in series with $C_p$) and using the series capacitance formula. $C'_{meas}$ is the capacitance predicted by (4), knowing C and $C_p$. Finally, $C_{meas}$ is the measured capacitance between the two top discs when the samples are placed on the grounded chuck. $C_{meas}$ and $C'_{meas}$ match to within 0.9%, thus illustrating the validity of Fig. 5 (b) and (4) as a model for the measurement setup.

## S-PARAMETER MEASUREMENTS

One-port S-parameter measurements are performed on the structure of Fig. 1 from 100kHz to 1GHz. A Hewlett Packard 8753E vector network analyzer (VNA) integrated with a Cascade probe station is used to carry out the measurements. Before measuring the test structures, a Short-Open-Load-Thru (SOLT) calibration is performed using Cascade Microtech's impedance standard substrate (ISS). The measured $S_{11}$ is illustrated in Fig. 8. To improve accuracy the data shown is a combination of measurements from individual sweeps over smaller frequency ranges.

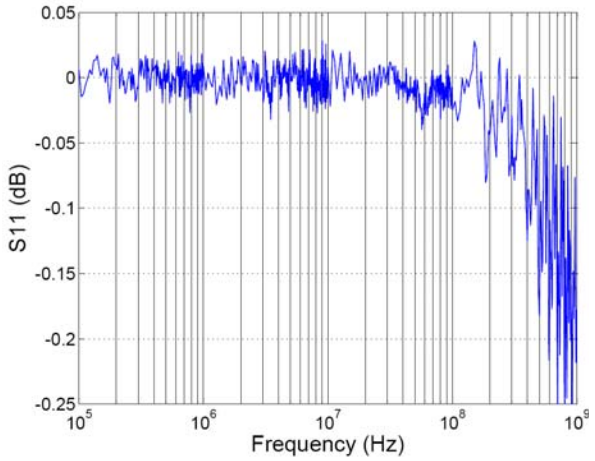



Fig. 8. $S_{11}$ from S-parameter measurements.

The measured $S_{11}$ is converted to the impedance of the device under test (DUT) by the following equations:

$$Z_{DUT} = Z_0 \frac{1+S_{11}}{1-S_{11}} \tag{5}$$

where $Z_0$ is the impedance of the measurement port of the VNA which is usually 50Ω. The capacitance value and loss tangent can be extracted from the impedance through the following equations:

$$C = \mathrm{Re}\left[\frac{1}{i\cdot\omega\cdot Z_{DUT}}\right]$$
$$\tan\delta = -\frac{\mathrm{Im}\left[\frac{1}{i\cdot\omega\cdot Z_{DUT}}\right]}{\mathrm{Re}\left[\frac{1}{i\cdot\omega\cdot Z_{DUT}}\right]} = -\frac{\mathrm{Re}\left[Z_{DUT}\right]}{\mathrm{Im}\left[Z_{DUT}\right]} \tag{6}$$

The calculated capacitance and the loss tangent are shown in Fig. 9 and Fig. 10, respectively.

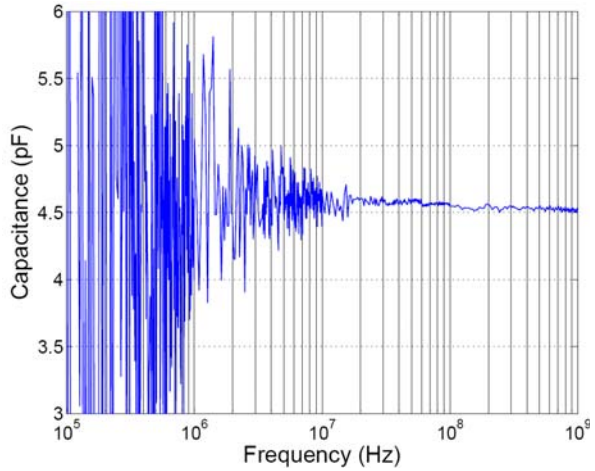



Fig. 9. Capacitance of a circular capacitor from S-parameter measurements.

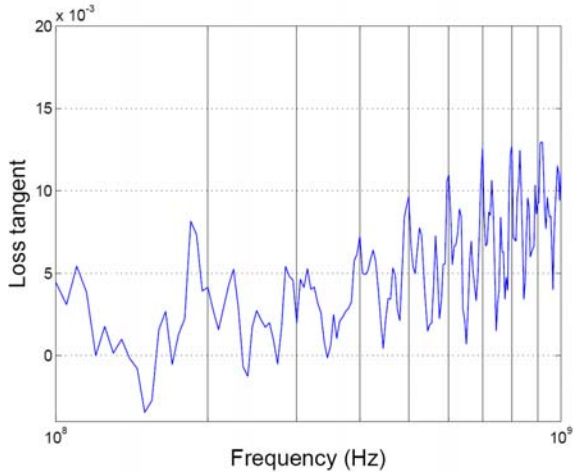



Fig. 10. Loss tangent from S-parameter measurements.

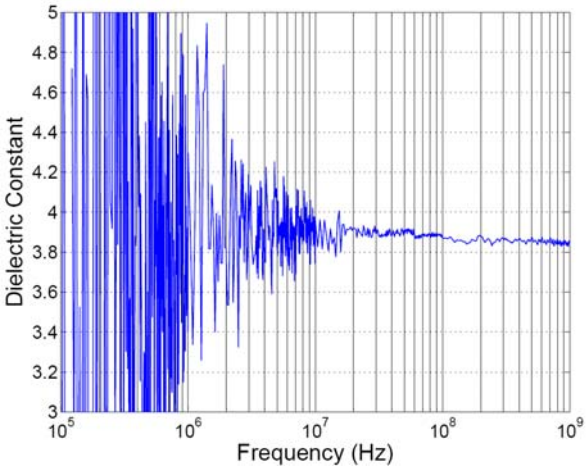



Fig. 11. Dielectric constant ($\varepsilon_r$) from S-parameter measurements.

Then, $\varepsilon_r$ can again be determined from (3). At high frequencies this is 3.85 which agrees within 1.3% with the value determined from the C-V measurements, thus giving further confidence in the measurement and extraction techniques.

## CONCLUSION

A technique has been developed to characterize thin film dielectric materials at RF frequencies. The technique requires only one lithography step on the top metal layer. Measurements are carried out with a vector network analyzer and C-V system, respectively. An analysis of the electrical properties of the test structure allowed the dielectric constant and loss tangent of the dielectric material to be extracted. The constancy between the C-V and S-parameter measurements gave increased confidence in the measurement methods, which can now be applied to high-k samples such as PMNT materials.

## ACKNOWLEDGMENT

This work is supported in part by the European Union Sixth Framework Programme through STRP Project 033103 (CAMELIA).

## REFERENCES


[1] J. A. O'Sullivan, W. Chen, K. G. McCarthy, G. M. Crean, "Advanced test structure design for dielectric characterisation of novel high-K materials," *Proceeding of the 2008 IEEE International Conference on Microelectronic Test Structures, ICMTS,* vol. 8, pp. 180-184, Edinburgh, 24-27 March 2008.

[2] Z. Ma, A. J. Becker, "RF measurement technique of characterizing thin dielectric films," IEEE Transaction on Microwave Theory and Techniques, vol. 45, no. 8, pp. 1811-1816, August 1998.

[3] M. P. J. Tiggelman, K. Reimann, J. Liu, M. Klee, W. Keur, R. Mauczock, J. Schmitz, and R. J. E. Hueting, "Identifying dielectric and resistive electrode losses in high-density capacitors at radio frequencies," *Proceeding of the 2008 IEEE International Conference on Microelectronic Test Structures, ICMTS,* vol. 8, pp. 190-195, Edinburgh, 24-27 March 2008.

[4] D. MacSweeney, K. G. McCarthy, L. Floyd, A. Mathewson, P. Hurley, J. A. Power, S. C. Kelly, "Influence of probing configuration and data set size for bipolar junction capacitance determination," *Proceeding of the 2002 IEEE International Conference on Microelectronic Test Structures, ICMTS,* vol. 15, pp. 127-132, Cork, 8-11 April 2002.

[5] Y. Okawal, H. Norimatsul, H. Suto, M. Takayanag, "The negative capacitance effect on the C-V measurement of ultra thin gate dielectrics induced by the stray caoacitance of the measurement system," *Proceeding of the 2003 IEEE International Conference on Microelectronic Test Structures, ICMTS,* pp. 197-202, Monterey, 17-20 March 2003.

[6] W. Chen, J. A. O'Sullivan and K. G. McCarthy, "Dielectric characterization for novel high-k thin films using microwave techniques," *Proceeding of IET China-Ireland International Conference on Information and Communications Technologies, CIICT,* pp. 29-36, Beijing, 26-28 September 2008.

# Test Structure Design, Extraction, and Impact Study of FEOL Capacitance Parameters in Advanced 45nm Technology

Shashank Ekbote, Priyamvada Sadagopan, Ying Chen, Wing Sy, Ron Zhang, and Michael Han
*QUALCOMM Incorporated, San Diego, CA 92121*
*Tel: (858)-658-3703, Fax: (858)-845-7307, e-mail: sekbote@qualcomm.com*

## Abstract

In the advanced Low Power (LP) CMOS technology nodes gate-to-source/drain overlap capacitance ($C_{OV}$), gate-to-contact capacitance ($C_{CO}$) and gate sidewall fringe capacitance ($C_f$) have become increasingly important component(s) of transistor parasitic. Accurate extraction and modeling of these parasitic are essential in accurate estimation of circuit performance. In this paper we describe test structure design and extraction of these parasitic components from silicon, which we later correlate to circuit performance. SPICE simulations were performed to substantiate the measurements as needed.

## Introduction

With the advancement in CMOS technologies, transistor sizes have been shrinking to achieve the increased packaging density. However, lightly doped drain junction (LDD) scaling have not kept up with the gate length scaling of the CMOS transistors. In addition, due to reliability and device performance concerns lateral diffusion (LD) or gate-to-drain/source overlap distance has not scaled proportional to transistor gate lengths. Fig. 1 shows the cross-sectional view of a MOSFET and the capacitances associated with the transistor, where gate length is the physical gate length of a self-aligned MOSFET, LD is the lateral diffusion of LDD region and $L_{EFFECTIVE}$ (Physical gate length – 2xLD) is the effective gate length of a transistor. Transistor capacitances are denoted by $C_{GATE}$ (gate plate to body capacitance), gate to LDD overlap capacitance $C_{OV}$, and gate to diffusion fringe component of the capacitance $C_f$. Gate-to-source/drain ($C_{GS}/C_{GD}$) capacitance is combination of overlap capacitance $C_{OV}$ and fringe component ($C_f$).

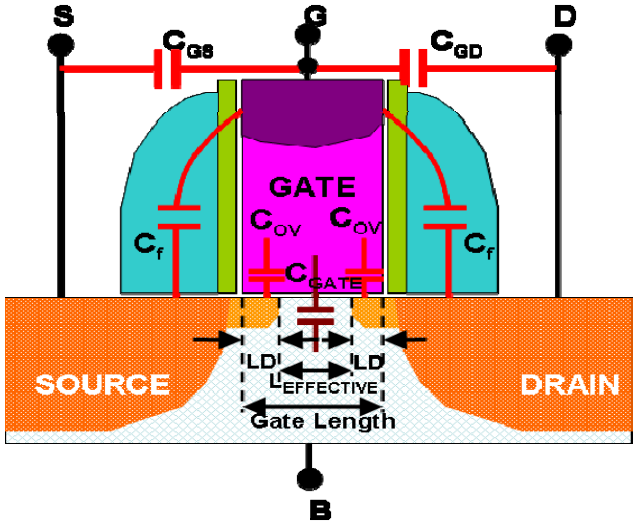



**Fig. 1** Cross sectional view of a MOSFET and associated gate capacitances

Inability to scale LD at the rate of gate length and gate dielectric scaling has resulted in $C_{OV}$ constituting larger part of total transistor capacitance ($C_{TOTAL}$) [1], [2]. Fig. 2 shows historical trends of $L_{EFFECTIVE}$ and LD scaling for Low Power (LP) Technology nodes. It is evident from the scaling trend that LD has not scaled proportional to $L_{EFFECTIVE}$ scaling. On the same figure secondary Y-axis represents $C_{OV}$ as a percent of total transistor capacitance. As expected, $C_{OV}$ comprises larger part of total transistor capacitance for recent technology nodes.

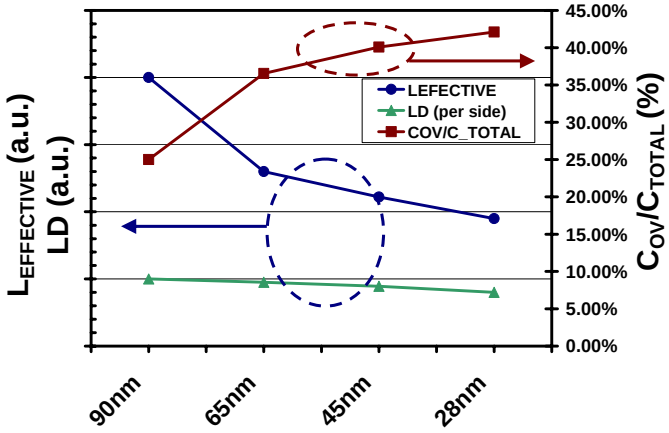



**Fig. 2** Contribution of $C_{OV}$ to total transistor capacitance for LP Technology nodes

It suffices to say that test structures design, extraction and modeling of parasitic capacitances ($C_{OV}$ and $C_f$) of a CMOS transistor has become extremely critical in evaluation of digital circuit performance.

### *Gate-to-contact capacitance*

In addition to transistor capacitances, gate-to-contact capacitance ($C_{CO}$) plays significant role in circuit performance and hence, in addition to optimizing intrinsic transistor capacitances (process optimization), contact density optimization (design optimization) is needed in order to meet circuit performance requirements.

In next sections we describe the test structure design, parasitic capacitance extraction and correlation to silicon data and SPICE simulation output. Finally, we correlate the performance of typical digital building blocks (such as inverter and NAND based ring oscillators) to the contact density, which is shown to correlate extremely well with contact density dependent Cf.

## Test Structure Design

As a first step to evaluate impact of transistor capacitances on circuit performance, design of test structures for extracting components of transistor capacitances is a key. On advanced 45nm LP CMOS technology node, we designed test structures for enabling extraction of key transistor capacitances. Fig. 3 shows a typical test structure design used to extract $C_{GS}/C_{GD}$ (i.e. $C_{OV}$ + $C_f$) and $C_{CO}$ components. $C_{GS}/C_{GD}$ are parasitic component of transistor capacitance whereas $C_{CO}$ is external capacitance modulated by various

design parameters such as contact to gate spacing, number of contacts etc..

Structures with varying number of contacts per side for a typical MOSFET structure were designed to extract combination of $C_{GS}/C_{GD}$ + $C_{CO}$ and then to separate out the $C_{GS}/C_{GD}$ component from $C_{CO}$ component. Structures were designed such that impact of Back-End-Of-Line (BEOL) RC parasitic was minimal on transistor parasitic extraction.

Fig. 3 shows test structure where number of contacts per side is varied. For fully contacted case, we expect the $C_f$ component to be minimal and $C_{OV}$ + $C_{CO}$ dominating. On the other hand, as number contacts per side reduces, contribution of $C_{GS}/C_{GD}$ (i.e. $C_{OV}$ + $C_f$) will increase and that of $C_{CO}$ will reduce.

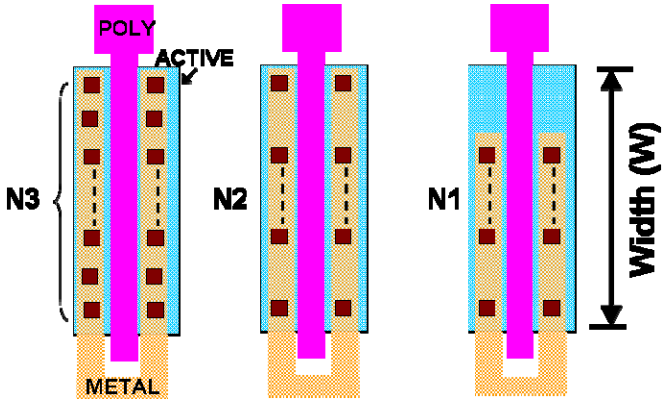



**Fig. 3** Test structure design for extraction of key capacitance Components of a MOSFET

**Capacitance Extraction and Correlation to SPICE**

In this section we describe the results obtained by characterizing three test structures with varying number of contacts (N1, N2 and N3) per side in a typical n-MOSFET.

In this section we present comparison of capacitances extracted from test structures (Data) vs. SPICE simulation results for two different cases; [a] $C_f$ (Const.), and [b] $C_f$ (accurate). Historically, fringe component of a transistor capacitance has been treated as a constant number per unit width. Recently, there have been efforts to model $C_f$ as a function of contact density and poly-to-contact spacing. In this study, we compare the silicon data with these two different cases [a] and [b]. As contact density (Fig. 3 Number of contacts N3-N1) reduces, more diffusion area is exposed and in case [b] $C_f$ changes with number of contacts, referred hereafter as $C_f$ (accurate). In sparsely populated contact case (N1), we expect $C_{CO}$ contribution to the total capacitance extracted to be minimal and its capacitance to be dominated by overlap ($C_{OV}$) and fringe component ($C_f$) of transistor capacitance. On the other hand, in densely populated contact case (N3), contribution of $C_{CO}$ will increase and that of fringe capacitance will reduce.

In order to validate our assumption, we extracted the transistor capacitances (from structures shown in Fig. 3) for 45nm LP core CMOS devices. Fig. 4 shows for a core N-MOSFET device comparison of silicon data vs. numbers extracted through SPICE simulation for cases [a] and [b] described above. This comparison clearly demonstrates that silicon data is in very good agreement with SPICE extracted numbers for case [b] where contact density dependence of $C_f$ is taken in to account. For case [a] where $C_f$ component of transistor capacitance is treated as a constant per unit width, there can be as much as 14% error in extraction of total transistor capacitance in case of sparsely contacted case (number of contacts N1).

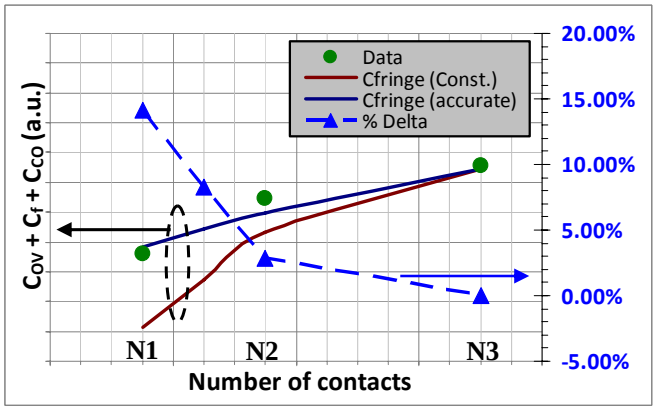



**Fig. 4** Key capacitive component ($C_{OV}$ + $C_f$ + $C_{CO}$) of a N-MOSFET as a function of number of contacts per side

Fig. 5 shows similar comparison for a P-MOSFET case. It suffices to state that both N- and P-MOSFET can have a large error (> 13%) in extraction of total transistor capacitances, especially for sparely contacted cases, if contact density dependence of fringe component ($C_f$) is not taken into account.

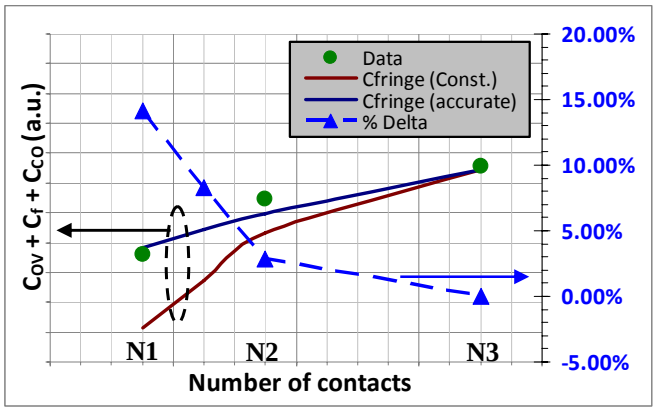



**Fig. 5** Key capacitive component ($C_{OV}$ + $C_f$ + $C_{CO}$) of a P-MOSFET as a function of number of contacts per side

After establishing importance of contact density dependent $C_f$ extraction, next logical step is to break-out the individual contribution of $C_{OV}$, $C_f$ and $C_{CO}$ to the total capacitance from these three key components ($C_{OV}$ + $C_f$ + $C_{CO}$). Fig. 6 shows as a function of contact density (number of contacts N1 through N3) contribution of individual components to the total capacitance ($C_{OV}$ + $C_f$ +$C_{CO}$) for cases [a] and [b]. As it can be seen from the graph, ignoring contact density dependence of $C_f$ underestimates its contribution to total capacitance by as much as 22% for sparsely contacted case.

In order to achieve node-on-node circuit performance improvement targets, it is mandatory to improve the device performances (such as DC saturation current $I_{DSAT}$) through process optimization, but, it is also extremely important to include the accurate estimation of ac parameters such as transistor capacitances [3] in circuit simulation. Special attention needs to be paid towards layout optimization, such as contact density, in order to reap to its full potential

benefits of a given technology node. In next sections we will focus on establishing correlation between performance of key Figure-of-Merit (FoM) digital building blocks, such as inverter and NAND based Ring Oscillator (R.O.) and transistor capacitances (through variation of contact density).

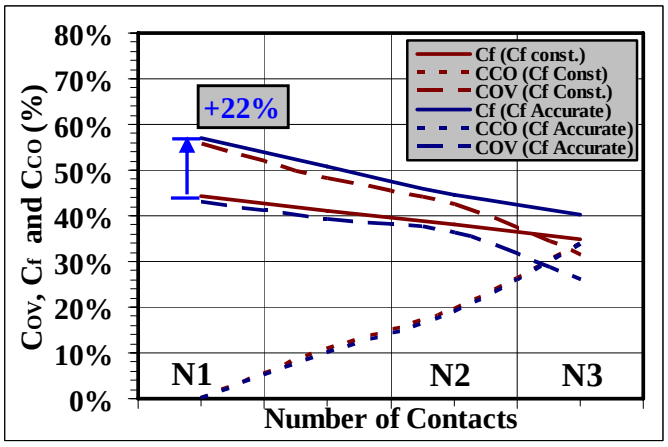



**Fig. 6** Individual contribution of $C_{OV}$, $C_f$ and $C_{CO}$ in % to total ($C_{OV}$ + $C_f$ + $C_{CO}$) capacitance for a N-MOSFET as a function of contact density

### Circuit Performance Correlation

In this section we will build-up on importance of contact density dependent fringe component of transistor capacitance when evaluating of circuit performance.

#### *Validation of Silicon vs. Model Output*

Before utilizing SPICE as a tool for evaluating impact of layout optimization (contact density in this case), it is very critical to validate the output of SPICE simulations against the silicon data. Fig. 7 shows a comparison of FO (Fan Out) =1 Inverter based R.O. delay data (silicon measurements) vs. SPICE simulated delay numbers for case [a] and [b]. This R.O. circuit has fully contacted transistors and was implemented in 45nm Low-Power CMOS technology node for FoM evaluation. Since both N-MOSFET and P-MOSFET devices in this case are fully contacted (maximum contact density), we expect contribution of $C_f$ to be minimal and transistor capacitance (excluding $C_{GATE}$) to be dominated by $C_{CO}$. We selected transistor dominated (gate dominated) R.O. circuit with minimized BEOL loading. Since $C_f$ is not a major contributor to transistor capacitance in this fully contacted case, both cases [a] i.e. constant fringe and case [b] i.e. contact density dependent $C_f$ extraction, show almost identical results as a function of effective $I_{DSAT}$ ($I_{DSAT,EFFECCTIVE}$). Silicon data shows excellent match with the SPICE extracted delay.

#### *Impact of Contact Density On Circuit Delay*

In above section, we validated the R.O. delay extraction methodology with silicon data and in previous sections; we also establish transistor capacitance correlation to contact density. Using cases [a] and [b] we can evaluate the impact of $C_f$ (fringe capacitance) component of transistor capacitance on circuit delay by varying the contact density. In case when contact density is a maximum, we expect R.O. delay from case [a] and [b] to be approximately identical and match the silicon data. As we reduce the contact density, case [a] i.e. assumption of constant fringe capacitance, will show sharper decrease in delay (for gate dominated R.O. case) due to reduction of transistor capacitance (as shown in Fig. 4 and 5). In contrast, case [b] which has accurate estimation of $C_f$, should show comparatively smaller changes in delay as contact density reduces (please refer to Fig. 4 and 5).

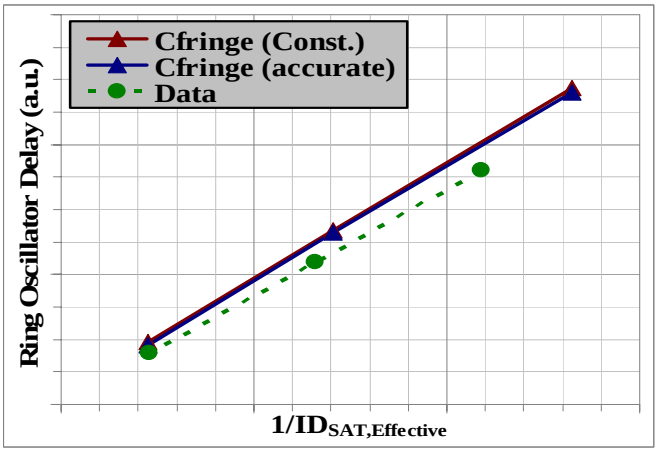



**Fig. 7** Inverter based (Fan Out = 1) R.O. delay vs. $1/I_{DSAT,EFFECTIVE}$ (comparison of silicon data vs. SPICE extracted numbers) for fully contacted transistor case (maximum contact density – referred to as C3 (ref.) here after)

In next sub-sections, inverter based (FO=1 and FO=3) and NAND based gate-dominated R.O. circuits are used as case studies to evaluate impact of contact density on R.O. delay.

#### *Case Study 1 – FO=1 Inverter based R.O.*

As a starting point we used inverted based FO=1 R.O. for above mentioned study. Fig. 8 shows comparison of R.O. delay as a function of contact density for cases [a] and [b]. Silicon data for C3 (ref.), which is fully contacted reference case, shows good matching with SPICE extracted R.O. delay for cases [a] and [b]. As expected, for fully contacted case, R.O. delay obtained for cases [a] and [b] are quite identical and matches very well with silicon data.

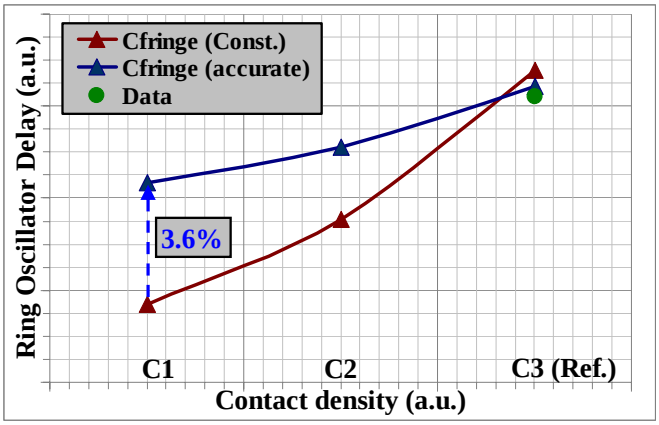



**Fig. 8** Inverter based R.O. (FO=1) delay as a function of contact density (where C3 (ref.) is fully contact case which was implemented and characterized on silicon and C3 > C2 > C1)

As contact density for transistors utilized in R.O. circuit reduces, we expect following two things to happen;

[1] Due to increased series resistance in current path of transistor (due to increased contact resistance), we expect some degradation in transistor saturation current ($I_{DSAT}$),
[2] As discussed in previous sections, with reduced contact density, transistor capacitance will reduce.

While decrease in $I_{DSAT}$ is expected to degrade the R.O. (i.e. increase in delay) performance, reduction of capacitance is expected to improve R.O. performance (i.e. reduction in delay).

As can seen from Fig. 9, as contact density reduces (C3 → C1) saturation current ($I_{DSAT}$) degrades (i.e. $1/I_{DSAT,\ EFFECTIVE}$ increases). Even though $I_{DSAT,EFFECTIVE}$ is degraded with reducing contact density, R.O. delay reduces (improves) which is due to the reduction of $C_f$ with reduced contact density. Unless contact density dependence of $C_f$ is included in the analysis (i.e. case [b]), impact of reduction of contact density on R.O. delay will be overestimated. As shown in Fig. 8 and Fig. 9, case [a] i.e. assumption of constant $C_f$ per unit width, overestimates reduction of R.O. delay, as a function of contact density, by as much as 3.6%.

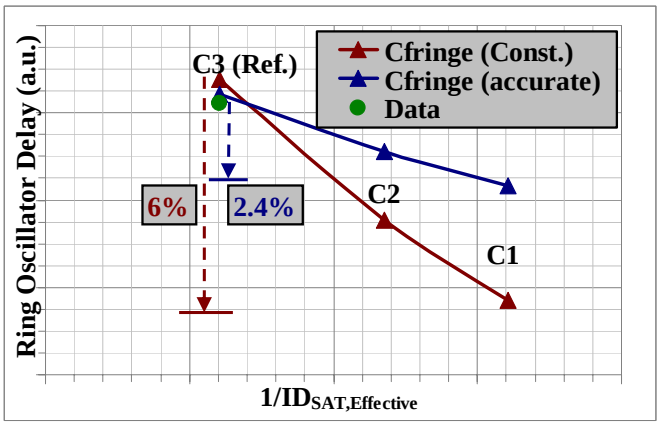



**Fig. 9** R.O. delay as a function of effective $I_{DSAT}$ for varying contact density (in case of FO=1,. Inverter based R.O.)

***Case Study 2 – FO=4 Inverter based R.O.***

In previous sub-section, we demonstrated that with reducing contact density on R.O. transistors, FO=1 inverter based R.O. delay can be underestimated by as much as 3.6% in case[a].

In this sub-section, we study the impact of contact density on FO=4 inverter based R.O. delay. Since R.O. that we selected for our studies are gate-dominated, we expect to have more impact of $C_f$ in case of FO=4 compared to FO=1 R.O. Fig. 10 shows comparison of R.O. delay vs. contact density for FO=4, inverter based R.O. Again, silicon data comparison to SPICE extracted delay numbers show good match validating our simulation results at fully contacted case.

Similar to the case of FO=1 inverter based R.O., again, it is evident that assumption of constant $C_f$ (independent of contact density) case shows sharper decrease in R.O. delay as contact density reduces. This can be correlated to reduction in $C_{CO}$ as a function of reduced contact density. However, when R.O. delay is extracted with accurate modeling of $C_f$, decrease in delay as a function of contact density is not as sharp as in case of constant fringe capacitance assumption.

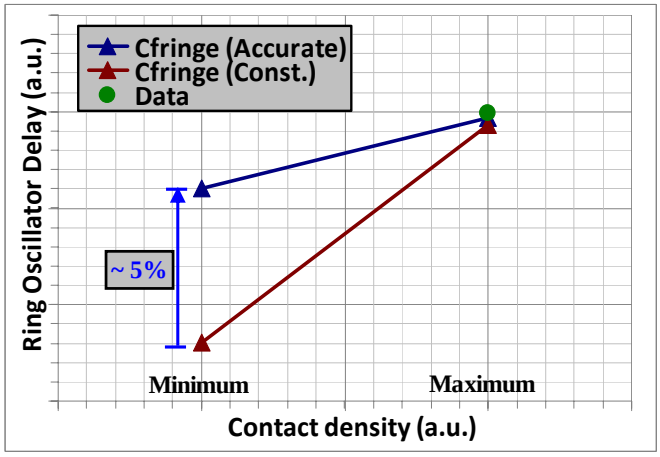



**Fig. 10** FO=4, inverter based R.O. delay as a function of contact density for case [a] i.e. Cfringe (constant) vs. case [b] i.e. Cfringe (accurate). Green solid circle represents median of silicon data for fully contacted R.O. case

Note that, in case of FO=4 inverter based R.O., error in estimation of R.O. delay can be as much as 5% if assumption of constant $C_f$ is made. Hence error in estimation of R.O. delay will decrease in case higher contact density is used for R.O. transistors, which due to increase in $C_{CO}$ will increase the R.O. delay.

***Case Study 3 – NAND based R.O.***

As a final case study, we used gate-dominated NAND based R.O. circuit. In case of inverter based R.O., width of N-MOSFET is approximately ½ that of width of P-MOSFET. In case of this specific NAND based R.O., width of P-MOSFET is same as that of inverter based R.O., but, since $W_P=W_N$ in case of NAND based R.O., contribution of N-MOSFET $C_f$ is expected to be higher in this case and we expect R.O. delay delta between case [a] and case [b] to be bigger.

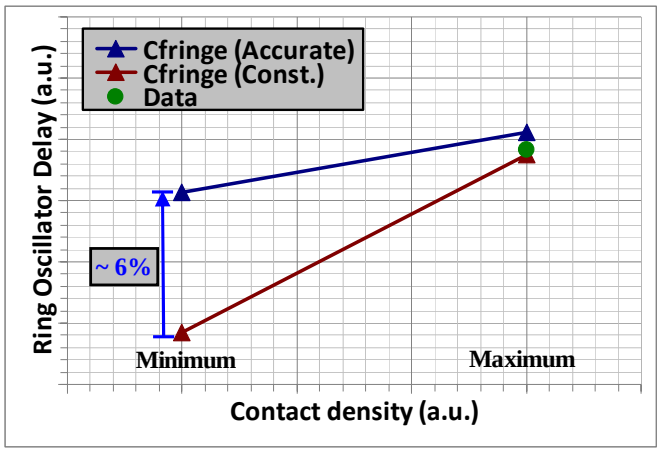



Fig. 11 R.O. delay vs. contact density in case of NAND based gate dominated R.O. circuit

As can been seen from Fig. 11, assuming contact density independent (i.e. constant) $C_f$ can lead to as much as 6% error in estimation of NAND based R.O. delay in this particular case. Again, validation of SPICE simulated numbers are done by comparing fully contact case with silicon data.

## Results and Discussions

While device performance improvement (e.g. saturation current, $I_{DSAT}$) through process optimization is required to improve the circuit delays, modeling tools need to account for accurate estimations of transistor parasitic. Error involved in extraction of transistor ac parasitic can lead to over prediction of circuit performance. We chose three specific cases of gate-dominated R.O. circuits with varying amount of gate loading (inverter based FO=1 and FO=4 R.O.) and with larger gate width (NAND based R.O.) to demonstrate impact of transistor ac parasitic on R.O. delay.

During state-of-the-art technology development, attention is typically paid to scaling (achieving packaging density), transistor performance improvement (process optimization techniques), BEOL RC optimization, but often transistor ac parasitic are ignored. By using gate-dominated R.O. in our case studies, we showed that transistor ac parasitic have become a critical part in estimation of circuit performance.

## Conclusion

In this paper we establish that transistor parasitic capacitances have become increasingly important contributor to circuit performance. While transistor geometries have scaled node-on-node to support higher packaging density, transistor parasitic capacitances are rather insensitive to geometrical scaling of transistor features. Accurate modeling of these parasitic have become important for advanced CMOS technology nodes to maintain and enhance predictive capability of modeling tools. Inaccuracies involved in extraction of transistor parasitic can lead to erroneous results and can over predict the circuit performance.

## References

[1] Jae-Rok Kahng, et. al, IEEE Int. Conference on Microelectronic Test Structures (ICMTS), Vol. 14, March 2001

[2] C.L. Huang, et. al, IEEE Trans. Electron Devices, Vol. 43, 1996

[3] Peter Bendix, IEEE Int. Conference on Microelectronic Test Structures (ICMTS), Vol. 2, March 1989