Impact of intra-die thermal variation on accurate MOSFET gate-length measurement

Ishtiaq Ahsan[1], Dieter. K. Schroder[2], Edward Nowak[1] , Oleg Gluschenkov[1], Noah Zamdmer[1], Ronald Logan[1]
[1]IBM Systems and Technology Group, 2070 Route 52, Zip 42J, Hopewell Junction, NY 12533 USA,
phone: (845)894-7627, email: ishtiaq@us.ibm.com.
[2]Arizona State University, Department of Electrical Engineering, Tempe, AZ 85281.

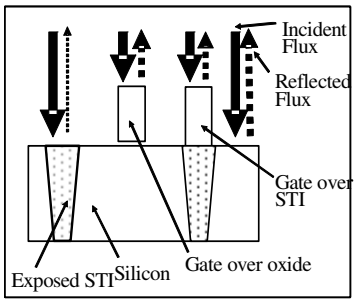
***Abstract* – It is known that significant intra-die thermal absorption variation is caused by non-optimized rapid thermal anneal (RTA) conditions and the variation depends on the local pattern density of various types of exposed stacks of the wafer. This variation can create errors in the electrical measurement MOSFET gate length itself. Two electrical methods for measuring gate length will be discussed, namely, the resistive technique, where a long-wide poly-silicon resistor is used as a normalizing resistor; and the capacitive technique, where a long-wide plate gate capacitor is used as a normalizing capacitor. It is shown, that the capacitive technique is more immune to errors introduced by RTA driven intra-die thermal absorption variation. Methods of minimizing these measurement errors are briefly discussed.**

I. INTRODUCTION

It is clear from recent work [1] that variation in device characteristics are becoming major yield limiters in deep submicron technologies. One of the key sources of device variability is the variability in gate-length since there is a very strong dependence of gate-length to many key device parameters like I_{ON}, V_T etc as seen in [2]. Hence it is very important to accurately measure the gate-length of the MOSFET. It has been shown that there is significant intra-die gate-length variation in sub-micron technologies [3]. The degree of intra-die variation described in [3] is sufficient to cause significant variation in MOSFET terminal characteristics like I_{ON}, threshold voltage (V_T) etc. In has been shown recently that significant intra-die thermal absorption variation is caused by non-optimized rapid thermal anneal (RTA) conditions**.** It will be shown that this variation can even create errors in the measurement of gate-length itself.

Intra-die thermal variation

Intra-die thermal variation is due to local reflectivity variation associated with different microstructures [4]. An example of one test-chip that demonstrated such variation will be discussed here [5]. Fig. 1 schematically illustrates the difference in reflectivity. The accompanying table lists the RTA-spectrum reflectance for different film stacks present on the wafer in the absence of scattering effects. Exposed shallow trench isolation (STI) stack had a markedly different reflectance in this case. For the test-chip under discussion, this difference created



Region	Reflection Coefficient
N+ Source Drain	0.57
P+ Source Drain	0.57
Gate over STI	0.54
Gate over oxide	0.45
Exposed STI	0.20

Fig. 1. Schematic and table of different reflection co-efficient of different exposed stacks in the absence of scattering effects [5].

significant variation of intra-die thermal absorption which resulted in intra-die variation in threshold voltage, ring-oscillator frequency etc. It is important to understand the test-structures and the test-chip of discussion to understand the nature of variation. The following section provides a description of the test-structures in the test-chip that was used. The diagram on the left in Fig. 2 is a schematic of the test-chip. Seven instances of one test-structure were placed on different locations of the chip (locations are denoted as "reference", UR, LR,LC,UC,UL and ML) and two instances of a separate test-structure (CAP1, CAP2) were also placed. The seven solid oval shapes indicate the placement of the first set of test-structures, each of which include devices to measure long-wide poly-silicon (poly-Si) sheet resistance (R_S) , resistance of minimum width (minimum gate-length) poly-Si lines, gate-capacitance of minimum gate-length features, MOSFET. parameters (e.g. V_T) and ring-oscillators frequency.

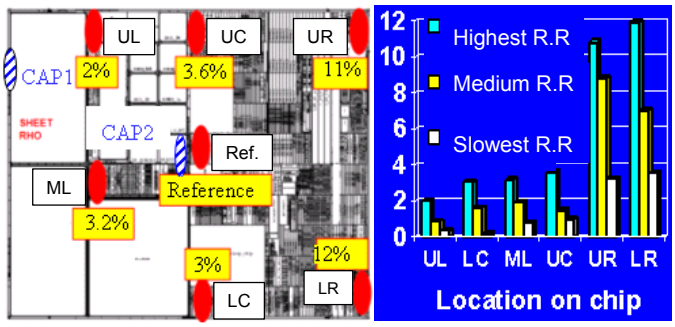


Fig. 2. Description of the test-structures in the test-chip and the degree of poly-silicon R_S variation due to thermal variation

The hashed blue oval shapes indicate the two placements of the 2nd type of test-structures that include devices to measure

gate-capacitance of a long-wide plate capacitor and the overlap capacitance. The numbers in yellow boxes indicate the amount of intra-die variation in the poly-Si R_s of a long-wide structure as measured from the reference location, indicating a significant intra-die R_s variation. The degree of this variation can be modulated by the ramp rate (R.R.) of the spike anneal process as indicated by the bar charts in Fig .2 (diagram on the right). This variation was shown to be driven by intra-die thermal absorption variation [5] which caused intra-die threshold voltage and external resistance variation, which in turn, caused intra-die ring oscillator frequency variation.

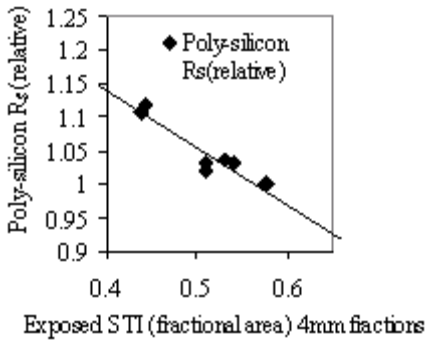


Fig. 3. Correlation between intra-die exposed STI pattern density and poly-silicon R_s variation [5]

Fig. 3 shows a good correlation between the poly-Si R_s (a thermal sensor) and the pattern density of "exposed STI" stack despite scattering effects caused by microstructures that are comparable or smaller than the wavelength in a typical RTA spectrum[5]. This finding allows for a simple correlation between RTA-sensitive electrical parameters and the identified pattern density.

Gate-length measurement

Fig. 4 illustrates the critical dimension of the MOSFET termed "L_{GATE}". This "bottom gate-length" of the MOSFET is the primary focus of this paper.

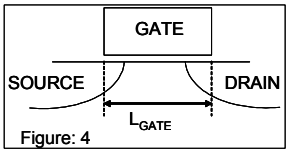


Fig. 4. Schematic of the L_{GATE} measurement of discussion in this work

In one popular traditional L_{GATE} measurement method referred to here as the "resistive technique"[6], (figure on right in Fig. 5), the resistance is measured from a "minimum L_{GATE}" poly-Si line from multiple locations on the chip and is normalized to the R_s measured from one long-wide poly-Si calibration resistor to extract L_{GATE}. It is now clear that this method is inherently prone to errors induced by the intra-die thermal absorption variation since the thermal absorption variation itself introduces variation in the R_s. For another popular L_{GATE} measurement method, referred to here as the "capacitive technique" [7] figure on left in Fig. 5); the gate-capacitance is measured for "minimum L_{GATE}" MOSFETs from multiple locations on the chip and is normalized to the long-wide "calibration" gate-capacitor to extract L_{GATE}.. This paper investigates if the "capacitive technique" is also impacted by the intra-die thermal absorption variation or not. This will be done by quantifying the intra-die variation of the calibration capacitor and comparing that to the variation of the calibration resistor.

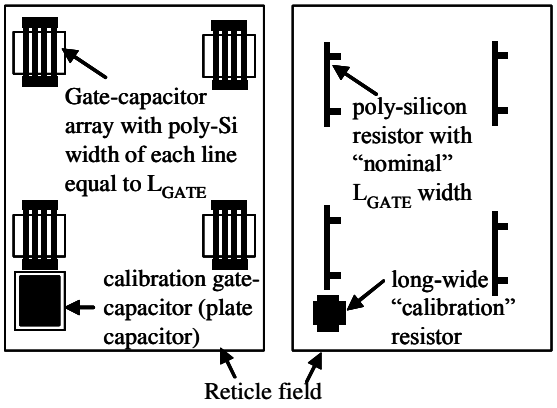


Fig. 5. "Resistive" and "capacitive" method of measuring L_{GATE}

II. EXPERIMENTAL RESULTS

The locations of the test-structures are shown in Fig. 2. It was not straightforward to estimate the amount of thermal absorption variation between the two locations were the CAP1 and CAP2 test structures were present since those two did not have the poly-silicon R_s monitors near them, However, based on the calculated exposed STI pattern density around the locations of the CAP1 and CAP2 test-structures, and using the correlation curve previously established in Fig. 4, one can project the expected poly-Si R_s difference between the CAP1 and CAP2 locations, as shown in Fig. 6. This figure indicates that had there been devices to measure poly-Si R_s in the CAP1 and CAP2 test-structures, there would have been a 7% difference in its value between the two locations.

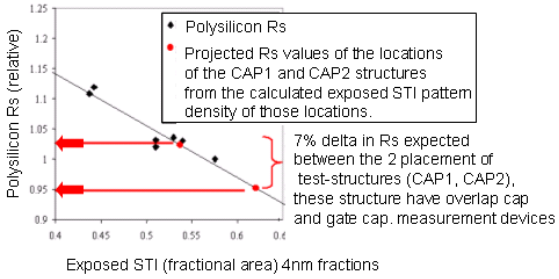


Fig. 6. Estimation of R_s delta between the CAP1 and CAP2 locations of the test chip.

Experiments were done on a lot were most of the wafers were processed with a non-optimized spike RTA with high

ramp rate. Two wafers of the lot were processed with a spike anneal process were the ramp rate was lower. Measurement was taken off the various test structures of the test chip described in Fig. 2. The results of some of these measurements are shown in Fig. 3.

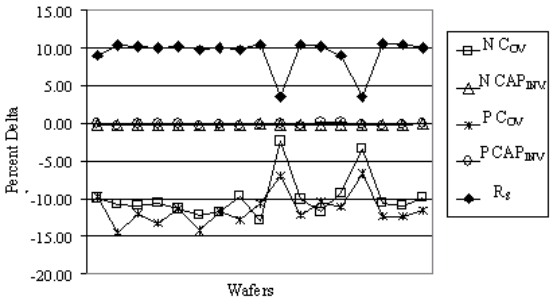


Fig. 7. Intra-die variation of different MOSFET parameters on an experimental lot, where two of the wafers were processed with a spike anneal with a lower ramp rate.

Fig. 7 shows the intra-die variation of various key parameters, overlap capacitance (N/P C_{OV}), inversion gate-capacitance (N/P CAP_{INV}) and poly-Si R_S (R_S). In this figure, the Y-axis shows , for all the parameters listed, the difference between the values measured from different locations of the test chip expressed in percentage terms. All the parameters except the R_S measurement come from the two placements of the macros CAP1 and CAP2. The R_S values come from the "reference" and "LR" locations as noted in Fig. 1. There is a significant variation (~10%) for the N/P C_{OV} between CAP1 and CAP2 in all but 2 of the wafers that were processed with the lower ramp RTA process. This indicates that for the high ramp RTA process, we would have measured a significant difference in the poly-Si R_S values between the CAP1 and CAP2 locations had those locations had the R_S measurement device. Fig. 7 indicates that this R_S delta is estimated to be ~7%. This estimation is further corroborated by another RTA temperature experimental lot which shows, as indicated in Fig. 8, that a ~15% difference in N C_{OV} driven by RTA temperature differences results in a ~12% difference in R_S (poly-Si R_{S2}). Hence it is reasonable to assume a ~7% difference in R_S for 10% NC_{OV} difference. One important observation from Fig. 7 is that there is very little intra-die variation in inversion gate capacitance between the CAP1 and CAP2 location despite having significant thermal absorption and poly-Si R_S variation. This is consistent with the RTA temperature experiment results (Fig. 8). In Fig.8, poly-Si R_{S2} was doped with half the doping concentration of a typical MOSFET gate poly, whereas poly-Si R_{S1} was doped with twice the concentration of a typical MOSFATE gate poly-Si. Both showed significant variation between the temperature experiments, whereas the gate capacitance variation was negligible.

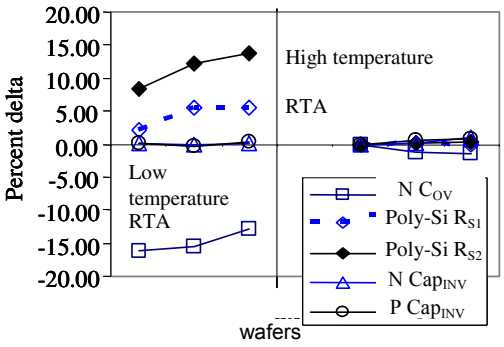


Fig. 8. Dependence of various MOSFET parameters on the temperature of the RTA process used. The data-points are wafer-median values of wafers run with different RTA temperatures with same ramp rate. All the points are normalized to the corresponding parameter values of the left-most wafer of high-temperature experiments.

III. SIMULATION RESULTS

A simulation of inversion gate-capacitance was done using a simulator named "SCHRED". A gate doping concentration of 10^{20} cm^{-3} and a gate-oxide thickness of 2.2nm were used. To emulate a 10% difference in poly-Si R_S due to thermal absorption driven dopant activation, the gate doping concentration was changed by 10% and inversion gate capacitance values were compared to the result of the original simulation. The simulated difference in inversion gate capacitance for a corresponding 10% difference in poly-Si Rs was only 1% (due to gate-depletion). Hence this first order simulation result is consistent with the observation regarding CAP_{INV} in Fig. 8 and 9, which is that thermal absorption driven poly-Si R_S variation is much more than the corresponding gate-capacitance variation.

To further demonstrate the minimal impact of the intra-die thermal absorption on the capacitive technique of L_{GATE} measurement, compact model simulation of ring oscillator frequency was performed for rings located in the "reference" and the "UR" location of the die. The MOSFETs in those locations had different threshold voltage and external resistance due to the difference in thermal absorption. Hence adjustments needed to be made for these parameters in the compact model for the simulations. The results were compared to a plot of L_{GATE} vs ring-oscillator frequency as shown in Fig. 10. The measured frequency is plotted against the measured L_{GATE} of the same location of the chip ("UR" or "reference"). The L_{GATE} measurement was done using the capacitive technique. The set of data points comes from different chips of the wafer. The variation in L_{GATE} comes from intentional, across wafer, lithography-induced variation. The fact that the model simulations match so well with the hardware data further validates the accuracy of the L_{GATE} measurement using the capacitive technique.

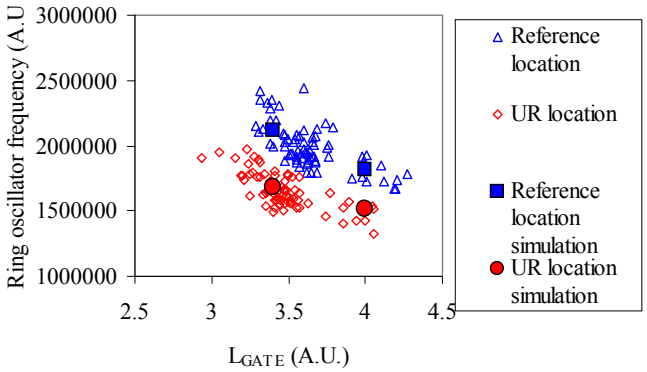


Fig. 9. Measured data and simulation: ring oscillator frequency vs L_{GATE}. The L_{GATE} measurement was done using the capacitive technique. The good agreement between simulation and measured data indicates consistency and accuracy in the measurement data of L_{GATE}.

IV. CONCLUSION

From all the discussion, it is clear now that the L_{GATE} measurement using capacitive technique is more immune to the intra-die thermal variation than the resistive technique because the gate capacitance itself is less susceptible to intra-die thermal variation than the poly-Si R_S. By placing separate calibration resistors near all the nominal line poly-Si test-structures and normalizing the nominal line poly-Si resistor to its local calibration resistor, this measurement error can be minimized. However, this comes at a cost of additional space on the chip. On the other hand, the capacitive technique is more costly in terms of test time. Ultimately the correct trade-off between test time and chip area needs to be made among other factors for the correct choice of L_{GATE} measurement technique.

ACKNOWLEDGMENT

This work was performed by the research alliance teams at various IBM research and development facilities. I would like to thank my colleagues in IBM, most notably Noah Zamdmer, for his assistance with the ring oscillator and hardware simulations and Oleg Gluschenkov for his help. I would also like to thank Ronald Logan, Bachir Dirahoi, Irene Popova of IBM and Hideki Kimura from SONY for their assistance. I would like to thank Ed. Maciejewski, Marybeth Nasr, James Rice, Matthew Paggi, Neil Peruffo, Paul Faraar, Ronald Geiger and, Xu Oyang and special thanks to David Riggs for management support in IBM.

REFERENCES

[1] M. Horowitz, E. Alon, D. Patil, S. Naffziger, R. Kumar, K. Bernstein, "Scaling, power, and the future of CMOS", *IEDM Technical Digest*, Section 1.2, pp. 7-14, 2005.

[2] Tyagi, C. Auth, P. Bai, G. Curello, H. Deshpande, S. Gannavaram, O. Golonzka, R. Heussner, R. James, C. Kenyon, S-H Lee, N. Lindert, M. Liu, R. Nagisetty, S. Natarajan, C. Parker, J. Sebastian, B. Sell, S. Sivakumar, A. St Amour, K. Tone, "An advanced low power, high performance, strained channel 65nm technology", *IEDM Technical Digest*, Paper 10.8, pp. 245-247, 2005.

[3] A. Mahorowala, S. Halle, A. Gabor, W. Chu, A. Barberet, D. Samuels, A. Abdo, L. Tsou, W. Yan, S. Iseda, K. Patel, B. Dirahoui, A. Nomura, I. Ahsan, F. Azam, G. Berg, A. Brendler, J. Zimmerman, T. Faure, "Meeting critical gate linewidth control needs at the 65 nm node", *Proceedings of SPIE*, Volume 6156, pp. 61560M, 2006.

[4] P. J. Timans, W. Lerch, J. Niess', S. Paul, N. Acharya', and Z. Nenyei, "Challenges for ultra-shallow junction formation technologies beyond the 90nm node" , *Proceedings of IEEE International Conference on Advanced Thermal Processing of Semiconductors*, pp. 17-33, 2003.

[5] I. Ahsan, N. Zamdmer, O. Glushchenkov, R. Logan, E. J. Nowak, H. Kimura, J. Zimmerman, G. Berg ,J.Herman, E.Maciejewski, A.Chan, A.Azuma, S.Deshpande, B. Dirahoui, G.Freeman, A. Gabor, M. Gribelyuk, S.Huang, M. Kumar, K. Miyamoto, D. Mocuta, A. Mahorowala, E.Leobandung, H.Utomo, B.Walsh, "RTA-driven intra-die variations in stage delay, and parametric sensitivities for 65nm technology", *Symposium on VLSI Technology Digest of Technical Papers*, Paper 21.2, pp. 214-215, 2006.

[6] D. K. Schroder, *Semiconductor Material and Device Characterization,* 2nd Edition, pp. 17, 1998.

[7] Y. Taur, "MOSFET channel length: extraction and interpretation," *IEEE Trans Electron Dev*, Vol. 47, No. 1, pp. 160-171, January 2000.