Double Patterning: Simulating a Variability Challenge for Advanced Transistors

Peter Evanschitzky, Alex Burenkov and Jürgen Lorenz
Fraunhofer Institute for Integrated Systems and Device Technology IISB
Schottkystrasse 10, 91058 Erlangen, Germany
peter.evanschitzky@iisb.fraunhofer.de

***Abstract*—In this paper a comprehensive study of the impact of variations resulting from double patterning lithography on SRAM performance is presented. In double patterning, feature sizes are reduced by splitting one mask level into two. Besides the increase of process complexity and costs a further penalty is the introduction of uncorrelated variations between the two incremental lithography steps employed.**

Keywords—Photo lithography; double patterning; SRAM; SOI CMOSFETs; 20nm gate length

I. INTRODUCTION

Process variations have become a major challenge for further device scaling and for advanced analog devices. Whereas some effects like Random Dopant Fluctuations are alleviated by advanced device architectures like FinFETs which employ undoped channels, novel process steps needed e.g. for the patterning for smaller feature sizes induce additional severe sources of variability. This paper deals with the impact of variations resulting from double patterning, which is used for optical lithography at 32 nm and below. The key problem is that in double patterning one mask layer is replaced by two masking steps. In turn, misalignment, defocus or other variations between these two steps lead to additional variations of devices and matching problems of otherwise identical components. In the following a 6T SRAM cell is used to demonstrate the double patterning specific scanner related variability aspects of focus and threshold variations.

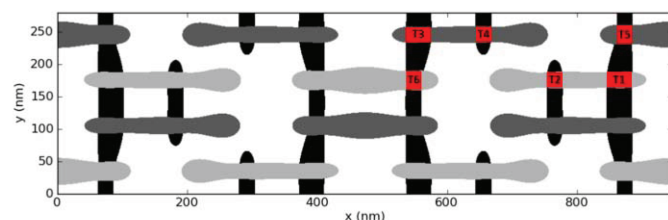

Fig. 1: Final result of the lithography simulation of the 6T SRAM cell exemplarily for one focus / threshold position. The black features are the active silicon areas, the dark gray features are the polysilicon areas generated with the first lithography step of the double patterning process and the light gray features are the polysilicon areas generated with the second lithography step of the double patterning process. Due to periodic boundary conditions four 6T SRAM cells have to be simulated. T1 / T2 / T3 / T4 are the flip-flop transistors, T5 / T6 are the access transistors. Due to the two lithography steps of the double patterning process the focus / threshold variations inside the two groups T3 / T4 / T5 and T6 / T2 / T1 are the same but between the two groups the variations are independent from each other.

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007 – 2013) under grant agreement no. 318458 SUPERTHEME

II. LITHOGRAPHY SIMULATIONS

In the lithography simulation part the structuring of active silicon areas and polysilicon gates of the 6T SRAM cell was simulated. Due to line widths down to 25 nm with pitches down to 70 nm a double patterning process had to be applied for the polysilicon layer whereas a single lithography step could be used for the structuring of the active layer [1]. For all simulations the Fraunhofer IISB tool Dr.LiTHO [2, 3] was used. The simulations are resist image based using a threshold model. In order to investigate the double patterning specific impact of focus and threshold variations each individual lithography simulation is divided into three basic parts: The computation of the mask diffraction spectrum, the resist image computation at all relevant focus positions and the evaluation of the resist images on the resist bottom at all relevant focus and threshold positions allowing a maximum variation of the Critical Dimensions (CD) of +/-15 % from the target. This procedure has to be applied to the three simulated lithography steps respectively: First lithography step of the polysilicon layer, second lithography step of the polysilicon layer and lithography of the active layer. The final result of the lithography simulation is shown in Fig. 1. Finally, undercutting of photoresist due to etching was taken into account by a fixed etch bias of 5 nm, to achieve the final physical gate lengths of 20 nm for the inner flip-flop transistors and 25 nm for the access transistors.

Due to the small lithography target feature sizes down to 25 nm of the presented SRAM example, the mask diffraction spectrum has to be computed rigorously using a rigorous electromagnetic field (EMF) solver. The resist image simulation, taking the mask diffraction spectrum as input data, is based on an extended Abbe approach. Instead of a full resist simulation a simplified resist image based threshold model was used for the CD computation of the different SRAM features. For all simulations the following lithography system was assumed: 193 nm water immersion with a numerical aperture of 1.35, a 4x reduction and a strong off-axis annular illumination with unpolarized light.

For an acceptable lithography process quality a simple Optical proximity Correction OPC had to be performed. In the first step, an attenuated phase shift mask with feature sizes corresponding to the target feature sizes of the respective

978-1-4673-5736-4/13/$31.00 ©2013 IEEE

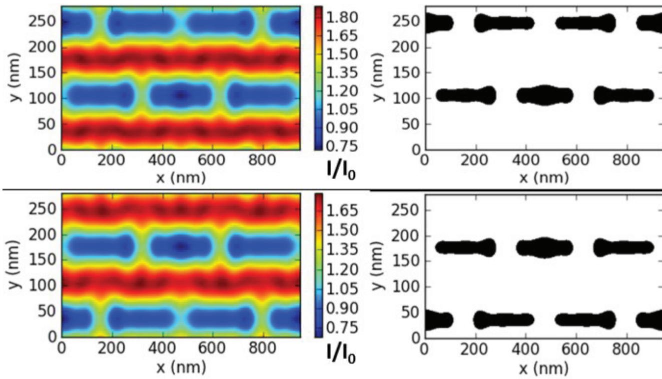

Fig. 2: Lithography of the polysilicon layer of the 6T SRAM cell. Resist images (left) and resulting feature profiles (right) on the resist bottom are shown exemplarily for one focus / threshold position of the first (upper) and second (lower) lithography step of the double patterning process. Structure generated in the first step are for clarity not shown in the lower right figure.

lithography step was used. In the following optimization loop the mask feature dimensions were modified at different positions in order to obtain the target sizes at the best focus position on the wafer side. This mask optimization procedure was used for the three simulated lithography steps.

For the patterning of the polysilicon layer a Litho-Freezing-Litho-Etch (LFLE) double patterning process was assumed. Due to the optical resolution limit of the described lithography system features with a pitch of 70 nm (see horizontal lines in Fig. 1) cannot be generated in one patterning step. Therefore, double patterning was applied. The mask was split-up into two masks with a resulting pitch of 140 nm on each individual mask. A first exposure with the first mask was performed. In the next step the photoresist was baked and developed. Proper thermal and/or chemical treatment renders the first resist inert with respect to a second exposure. After that, a second photoresist was spin coated and exposed with the second mask. The second exposure is affected by the wafer topography resulting from the first exposure. This wafer topography includes a slight modification of the optical properties of the frozen resist after the first lithography step. Similar to the well-known rigorous EMF effects induced by the mask topography, the wafer topography inside the wafer stack can result in topography effects such as reflective notching,

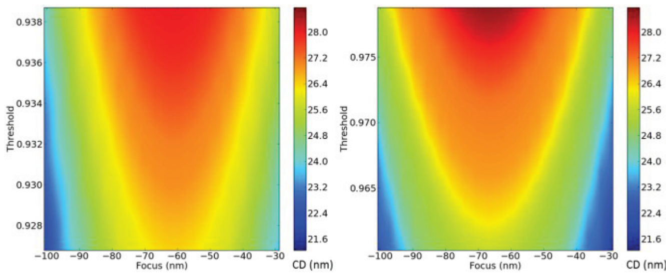

Fig. 3: CD (gate length) distributions depending on focus / threshold variations with a maximum CD deviation of +/-15 % from the target (25 nm). The pictures show the gate lengths of the transistors T2 (left) and T4 (right). The transistors are shown in Fig. 1. The corresponding polysilicon lines are generated with the described double patterning process (see text) in two independent steps: T4 with the first lithography step, T2 with the second lithography step. The differences of the distributions can be seen clearly. Furthermore the threshold levels of the two lithography steps are different (see threshold axis in the pictures). Due to the two lithography steps the focus / threshold variations of T2 and T4 are independent from each other.

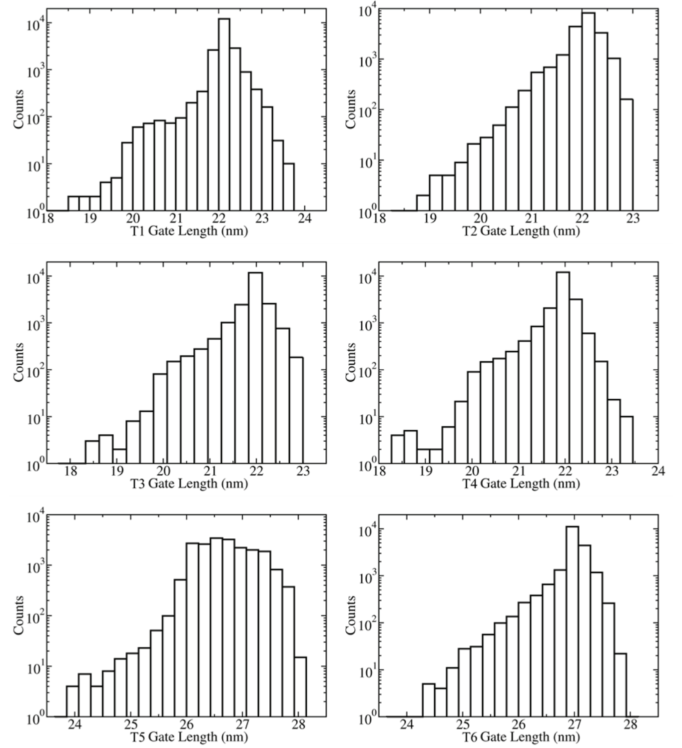

Fig. 4: Gate length variations of the six transistors T1 to T6 of the SRAM cell calculated under the assumption of Gaussian variations of the defocus and of the intensity threshold.

resist footing, reduced efficiency of the bottom antireflective coating, CD variations and other exposure artifacts. In order to cover all these effects, rigorous EMF simulations have to be applied to the computation of light diffraction inside the wafer stack. Due to the complex illumination of the wafer stack caused by the imaging of a mask with a projection system, the EMF wafer stack simulations are more challenging compared to EMF mask simulations. In order to realize acceptable simulation times, in the lithography simulator Dr.LiTHO [2, 3] of IISB optimizations and extensions of the rigorous EMF solver and of the simulation flow including all parts of the lithography system have been implemented. Fig. 2 shows the corresponding simulations of the two lithography steps for the patterning of the polysilicon layer. Among other things, two important differences of the double patterning process compared to single patterning have to be taken into account. The structures generated by the second lithography step (see lower line of Fig. 2 and light gray features of Fig. 1) are impacted by the wafer topography resulting from the first lithography step. Therefore, the CD distributions resulting from focus / threshold variations of the second lithography are different compared to that of the first lithography. Furthermore the focus / threshold variations of the two steps are independent from each other. Fig. 3 demonstrates this situation exemplarily for two transistors.

III. VARIABILITY OF CRITICAL DIMENSIONS

In the result of the photo lithography simulations, process windows for the three lithography levels presented before were identified. The process window defines the tolerable range of defocus and of the intensity threshold for each of the

lithography levels. Each of the three lithography levels has its own process window.

To study statistical variations of the critical dimensions we assumed that defocus and intensity threshold vary in each lithography step independently, having a Gaussian distribution inside of the process window. The standard deviation of this distribution was assumed to be 1/8th of the maximum tolerable variation range in the corresponding process window. The resulting statistical distributions of the gate length and of channel widths are presented in Figs. 4 and 5. The distributions of the gate lengths and of channel widths are far from being Gaussian and a specific distribution of these critical device dimensions is observed for each transistor in the SRAM cell. The distribution of the final critical dimensions would be Gaussian, if the critical dimensions shown exemplarily in Fig. 3 would have a linear dependence on defocus and intensity threshold. But in fact the CD dependence on defocus and intensity threshold is strongly non-linear and this non-linearity is different for each photo lithography level and even for each transistor in the SRAM cell. In result, instead of the Gaussian distributions of CD we see in some cases in Figs 4 and 5 sharp maxima and in other cases flat maxima. An asymmetry and in some cases an appearance of additional bumps are typical features of the resulting CD distributions. A bump or local maximum results in those cases when there are some areas in Fig. 3 with a weak dependence of CD on defocus or intensity threshold. From the asymmetric distributions of the critical device dimensions also asymmetrical distributions of the performance parameters of the SRAM may be expected.

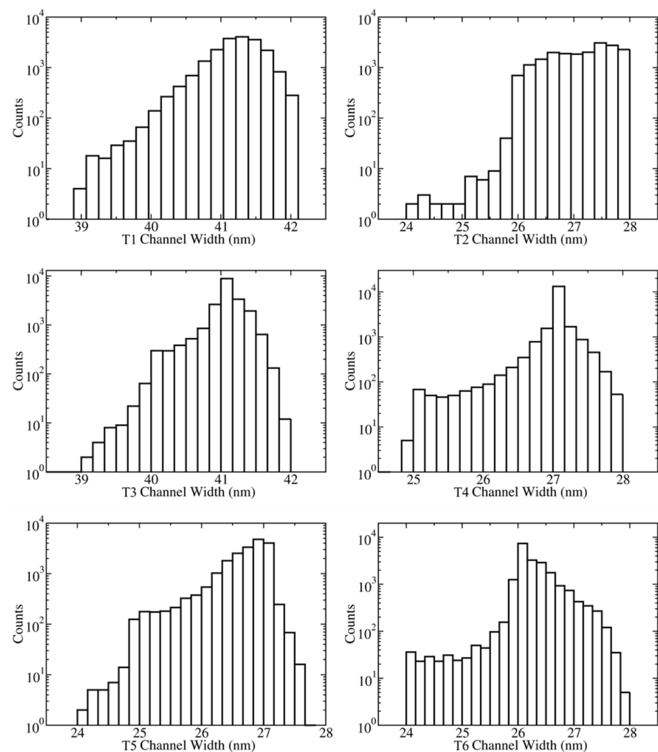

Fig. 5: Gate width variations of the six transistors T1 to T6 of the SRAM cell calculated under assumption of Gaussian variations of defocus and of intensity threshold.

IV. Variability of electrical performance of SRAM

The electrical performance of the SRAM circuit was simulated using the HSPICE simulator of Synopsys. The compact model BSIMSOI4 used in the HSPICE simulation was extracted from the results of Sentaurus TCAD simulations [4] for the fully-depleted SOI CMOSFETs with gate lengths in the range from 13 to 40 nm. For the statistical analysis of the SRAM, transistor CD distributions were extracted from the photo lithography simulations presented above, assuming random values of defocus and of the intensity threshold, with Gaussian distribution as described in the previous section. These values of the gate length and of the channel width were then transferred to the input decks for HSPICE and in this manner about 20000 simulations of the SRAM circuit were performed for the statistical analysis of the SRAM performance, similarly to our previous work [5] for an SRAM of the 32 nm gate length CMOS technology generation.

Prior the full scale statistical analysis, we performed a simple sensitivity analysis to observe what the expected variability of the performance parameters is and if the correlation between the variations at different photo lithography levels has an impact on the performance parameters. For this purpose we first set defocus and intensity threshold to their minimum and maximum values and observed the resulting change of the SRAM performance parameters. Fig. 6 shows the results of such sensitivity analysis for two variants of correlated variations. In the first variant shown in Fig. 6 left, defocus and intensity threshold were set simultaneously in all lithography levels to their minimum and then to their maximum values. In Fig. 6 right, defocus and intensity threshold were set in anti-phase in the two lithography levels for polysilicon structuring, i. e. in Var F1 defocus was at its minimum in the first illumination but at its maximum in the second illumination; in Var F2 defocus was at its maximum in the first illumination but at its minimum in the second illumination; in Var D1 the intensity threshold was at its minimum in the first illumination but at its maximum in the second illumination; and in Var D2 the intensity threshold was at its maximum in the first illumination but at its minimum in the second illumination. The results of such sensitivity analysis are shown in Fig. 6. If defocus is set synchronously to its minimum a lower Read static noise margin than in nominal setting is obtained. On the other hand, if defocus in Var F1 was at its minimum in the first illumination but at its maximum in the second illumination, we have very little difference in Read SNM in comparison to the optimum setting of defocus and intensity threshold in the middle of the process window (Nominal). And in contrast, in Var F2, when defocus was at its maximum in the first illumination but at its minimum in the

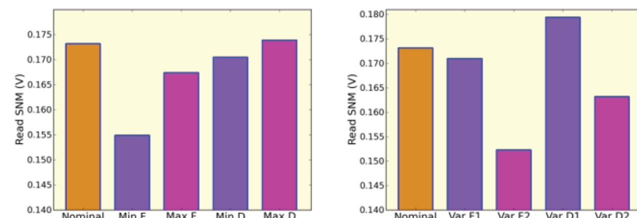

Fig. 6: Read Static Noise Margin variations of the SRAM circuit from minimum/maximum sensitivity analysis.

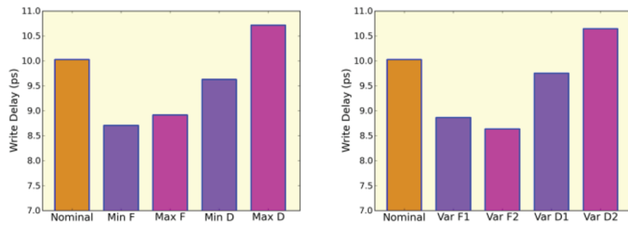

Fig. 7: Variations of Write delay times of the SRAM circuit from the minimum/maximum sensitivity analysis.

second illumination the value of the Read SNM has a value much lower than nominal SNM or SNM in Var F1. A synchronous variation of the intensity threshold leads to moderate change of Read SNM, while Var D1 variation results in a SNM significantly higher than the nominal and Var D2 leads to SNM significantly lower than the nominal.

A result of the sensitivity analysis for the Write delay time is shown in Fig. 7. Variations of defocus and intensity threshold are the same as those presented in Fig. 6. The sensitivity of the write delay to those variations is different from the sensitivity of the SNM margins, but also some common features in the impact of the defocus and intensity threshold variations can be seen.

Summing up, we conclude that the resulting impact of defocus and intensity threshold variations on SRAM performance depends not only on the magnitude of the variations but also on the correlations between those variations in separate illumination levels. The impact of variations from the first and second illumination in double patterning photo lithography is significantly different.

Figs. 8 and 9 present some of the results of the statistical SRAM performance analysis. Statistical distributions of Static Noise Margins (SNM) for the Read and Write mode of SRAM are shown in Fig. 8. The distribution of the SNM for the Read cycle is asymmetrical, left and right sides of the distribution look rather different. A rather long tail of the distribution to smaller values of the Read SNM is critical for SRAM performance. The distribution of the Write SNM values shown

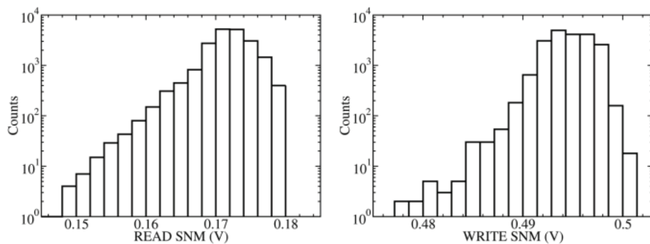

Fig. 8: Distributions of Read (left) and Write (right) Static Noise Margins (SNM) of SRAM circuit extracted from the statistical SRAM performance analysis.

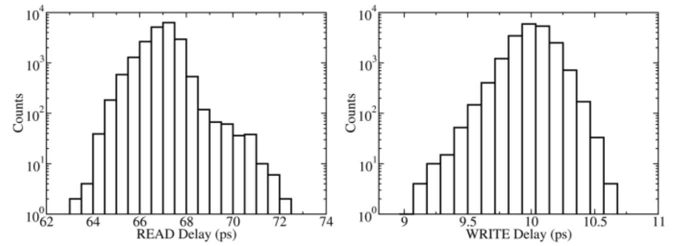

Fig. 9: Distributions of Read (left) and Write (right) delays of SRAM circuit extracted from the statistical SRAM performance analysis.

in Fig. 8 right has a similar shape but a relatively wider maximum and the relative spreading of the values is low considering the higher values of the Write SNM in comparison to the Read SNM.

Fig. 9 presents the results of the statistical analysis of the delay times for Read and Write cycles. The read-delay distribution has a longer tail to long Read-times. This may be critical for SRAM performance. The distribution cannot be approximated well by a Gaussian distribution behind the maximum. The Write cycle delay distribution has a more regular but asymmetrical shape. The longer tail of the distribution extends to shorter Write-times. This extension to shorter Write-times is not dangerous. The extension of the Write-delay distribution to longer delay-times is critical, but the shape of the distribution behind the maximum can be approximated with a half-Gaussian distribution and the probability of long delay-times can then be well predicted.

V. CONCLUSION

This paper complements the simulation approach for the impact of process variations presented in [1] and demonstrates that the introduction of double patterning introduces a new and challenging source of variability. Especially, variations resulting from the two incremental lithography steps employed are not correlated and therefore lead to unexpected impacts on SRAM performance.

REFERENCES

[1] A. Miller, J. Provoost, M. Maenhoudt, Double patterning lithography: Reducing the cost, Future Fab International, 2009, Issue 29, April 2009, pp. 66.

[2] T. Fühner, T. Schnattinger, G. Ardelean, A. Erdmann, “Dr.LiTHO – a development and research lithography simulator”, Proc. SPIE 6520, 65203F-1, 2007.

[3] www.drlitho.com

[4] Sentaurus TCAD, Release H2013.03, Synopsys Inc., Mountain View, CA, USA, 2013.

[5] A. Burenkov, C. Kampen, E. Bär. J. Lorenz, „Correlation-aware analysis of the impact of variations on circuit behavior“, Proc. SISPAD 2012, pp. 260-263.