Metrology Limits of Mask Process Development

Pavel Nesladek, Andreas Wiswesser, Björn Sass, Jan Richter
Advanced Mask Technology Center, Rähnitzer Allee 9; 01109 Dresden, Germany

ABSTRACT

The ever-narrowing specifications for high-end masks can only be derived from the continuous improvement of all manufacturing processes. Here, the metrology is crucial prerequisite since the development relies almost entirely on measurement results. In this paper we will address this relation by showing how the limits of metrology repeatability and reproducibility define also the limits of process development. In particular, we will show that improved metrology tool performance on resist results in a deeper understanding for the dry etch process. This is very important since resist metrology is not part of the ITRS roadmap and serves "only" as a supporting engineering process.
Better short-term repeatability results in the possibility to detect more variables that might influence the etch regime. As an example, results from two CD scanning electron microscopes (SEM) were compared with very different short-term repeatability. The better knowledge based on the more accurate metrology data allows then to optimize the process within a process space which was previously not detectable with the other tool. An estimate is given how much this influenced the final performance of the process. We conclude from these results, that metrology parameters not covered in standard roadmaps become increasingly important to achieve process development goals in other process areas.

Keywords: Metrology, process optimization, etch process, design of experiment

INTRODUCTION

The importance of accurate metrology measurements has been discussed in countless publications and its central role has been underlined in numerous specifications in the ITRS roadmap. Usually the ITRS roadmap specifications target at the final product and mostly derive their numbers from these requirements. In this paper we want to extend the discussion of the measurement tool influence to the single process and would like to investigate the role of metrology for process development that occurs prior to the final product assembly. Here, the budget model is a common tool to meet the final product specifications on the unit process level that basically divides the specification in smaller parts and assigns one of these parts to each contributing process. Each process is then developed to meet his internal budget. This brings us to the important observation that for internal customers like process developers the metrology specifications in principle should be downscaled according to their budget of the final product. As a matter of fact in most cases these downscaled requirements can not be met by metrology tools and in some cases there is even not a specification requirement in ITRS at all. These considerations raise the question if it is beneficial to utilize tools that are well below the specifications of the ITRS roadmap and are more in line with the invisible needs of internal customers. This implies that a better (than required by ITRS) metrology tool would lead to a better process development.

In our work we will develop means to understand errors in etch process models due to statistical measurement errors and their consequences. These models are then used for process development and represent the main method to reach the internal process requirements. In etch development for example the uncertainty in resist thickness measurement impacts directly the precision of the resist etch rate. However, there are countless not so obvious relations. In any case the precision of the resist etch rate estimation determines which influencing factors will be "visible" to the experimenters. Factors showing strong effect will be more visible than factors with weaker effects.

In some cases one can shift the resolution limit in order to see weaker effects, for example by increasing the experiment window or introducing repetition into the Design of experiment (DoE). Unfortunately, in many cases even the limits are unknown and thus the weaker factors can not be optimized or in the worst case the measurement error is misinterpreted as an effect. Finally, the result is an imperfect model of the process within the investigated process window which can in extreme circumstances be also misleading.

In this work an imaginary etch process is used to evaluate the impact of different metrology precision. The metrology is used during process development many times: firstly, for the estimation of the border conditions of the experiments, secondly, for the estimation of the effect of the process, and thirdly for the effect of a particular factor within the process. In each of these different steps the requirements on metrology are a little different with the second point being the most crucial for developing the correct model.

Therefore, we will focus in this paper on the estimation of the effects of the process during the DoE experiment, when the experiment window and conditions are adjusted already. The goal of the metrology in this phase is to provide an exact picture of the dependencies influencing the process and determining the parameter of the final product.
Nowadays statistical methods like Design of experiment are widely used for process optimization, providing very detailed information about the process behavior within the experiment window based on very limited number of runs.[1] If this methodology is applied correctly, this approach opens up a broad space for rapid process development and optimization and at the same time reduces costs.
Nevertheless we have to keep in mind limiting factors and assumptions being made and clearly mark the validity range for the conclusion drawn of the data obtained.

- The statistical methods provide no physical model of the nature so the conclusions are valid only within the range of experiment conditions used with very limited prediction capability outside of these limits.
- The precision of the model cannot be improved by any kind of regression or fit to some valid physical model and it is strongly dependent on the error within the experiment set up and the error in the estimation of the responses of the experiment.
- The statistical model assumes normal distribution of the error function and any deviation from this error distribution may lead to misinterpretation of the data obtained.

As a simple example we may perform the Galileo's experiment. An iron ball rolls along a ramp and we measure the distance from the starting point as function of time. In contrary to Galileo we will measure the distance at 2 selected times $\mathbf{t_1}$ and $\mathbf{t_2}$ only. Comparing both experiments we get the following diagram illustrating the difference between statistical and physical model.

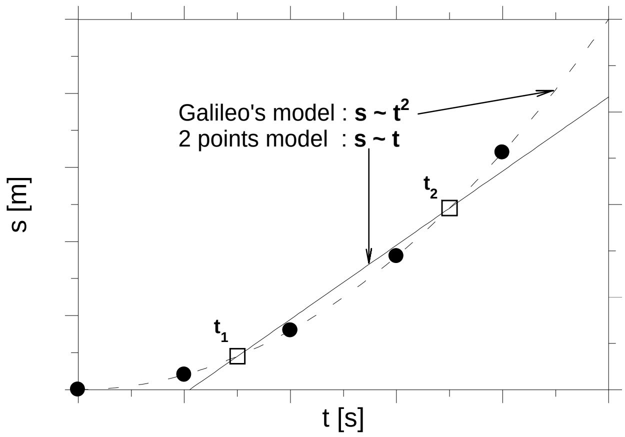


Fig.1 - Galileo's experiment compared to 2-point statistical experiment set up.

The design of the "statistical" experiment (see Fig.2) with 2 experimental points and the conclusion are obviously wrong compared with 3rd person having more observations available or knowing the right formula. This error is very frequently done, when observing more complex processes with less chance to easily obtain more data points.
As said, the 2-point model is physically wrong, however within the time range $\langle \mathbf{t_1}; \mathbf{t_2} \rangle$ fits quite well with the Galileo's physical model. With certain precision abandonment, one gains an acceptable model for the investigated range. In both experiments the precision of experimental work and the estimation of time and distance play a comparable role.
To estimate the precision of the statistical model as function of the measurement error for several different experiments is the goal of this paper.

METHODOLOGY

A "nature model" for a 4 factor full factorial design with 2 levels and 3 center points is established by the set up of the coefficients for each factor A, B, C … and calculation of the values for the 19 runs in the corresponding set of DoE data. A "statistical" model is calculated by adding an "error set", i.e. a group of normally distributed values with mean = 0 and predefined sample standard deviation σ to the "nature model". To generate the normally distributed error sets the

Mersenne-Twister[2] and inversion method[5] random number generator were employed. That way we obtain “statistical DoE sample” data and evaluate these as any other DoE data using Analysis of variance (ANOVA) method. Variation of σ of the error set allows us to estimate quite precisely the effect of the random error on the model. Directly from the ANOVA analysis we obtain the number of factors in the statistical model and the effect of each of the significant factors. The number of the factors found can be used as criteria for the statistical model obtained. However the critical point is the deviation of the statistical model from the nature model at any conditions within the DoE process window. We decided to judge the statistical models on hand of the maximum and average deviation between both models, calculated for all 19 run conditions (16 vertices and 3 center points) and call the values obtained average and maximum model error.

During the data evaluation we have seen, that the error of the model is not only dependent on σ of the error set, but shows additional variation, which will be investigated separately. Following scenarios were examined:

- influence of σ of the applied error set on the average and maximum model error
- effect of variation of σ estimated at center points on the average and maximum model error
- effect of the randomization/sorting of the runs within the DoE set

These effects were estimated for single measurement without replication. We are aware, that using the metrology twice, before and after the process investigated, to estimate the contribution of the process and/or the spatial distribution of the measurement for estimation e.g. critical dimension distribution (CDU) changes dramatically these effects. Since the estimation of the effect in all applicable cases is quite difficult and has to be estimated for each case separately, we focus this time on the case mentioned before. Other cases will be solved more detailed subsequently.

All investigations were done using 4 factors full factorial design with 3 center points with total number of runs equal 19 as mentioned before, which represents one of the simple and mostly used designs.

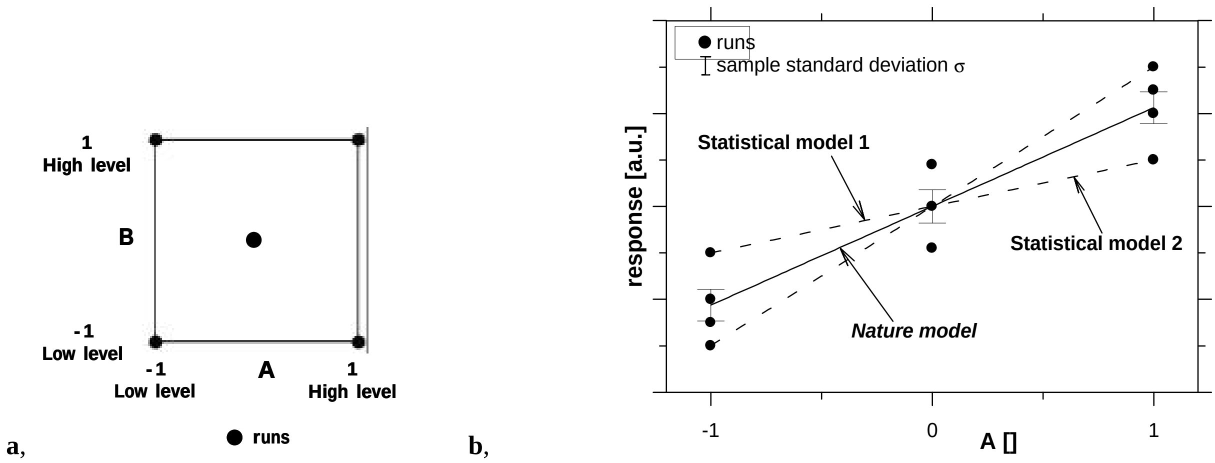


Fig.2 - **a**, Example for design of experiment with 2 factors **A** and **B** with 2 levels and center point **b**, effect and sample standard deviation for effect **A** as estimated from the experiment runs. The statistical model 1 and statistical model 2 show the maximum variation of the response.

Experimentally performed observation can be described as

$$m = m_0 + e_m + e_x \qquad [1]$$

- where m is the observed value; m_0 the real value corresponding to planned experimental conditions; e_x is the deviation due to variation in experiment and e_m is the measurement error. In our investigation the e_x term is omitted since there is no process deviation included in our model. When pre and post measurement is involved, the formula can be expressed as:

$$m = m_0 + e_{m1} + e_{m2} \qquad [2]$$

- where e_{m1} and e_{m2} is the error of the pre and post measurement respectively.

ERROR OF MEASUREMENT

As mentioned in the introduction, the simplest experiment we investigate is single measurement like film thickness measurement after the coating process.
For evaluation purpose an imaginary process was established involving all 4 factors of the 4 factorial design. Since the effect of each factor varies (see Table 1) we expect different sensitivity of the factors depending on the error set. From the original process model (We will call it "nature model" to avoid confusion with the models estimated from data sets) the values for each run were calculated.

Every single analysis was based on probabilities $\alpha = 0.05$ and $\beta = 0.10$. Factors were selected by additive method so that next factor was added until the probability of the new selected factor was > α or a three-factor interaction was selected. The last selected factor was removed again and remaining factors build the model for which the effects were calculated.
E.g. let's assume following factors and their probability for a 4 factor DoE:
A - 0.001; **B** - 0.005; **C** - 0.007; **AB** - 0.010; **ABC** - 0.040; **BC** - 0.045
The factors remaining in the model will be A, B, C, and AB interaction. Even so the probability for interaction BC is below 0.05 it will be removed from the model.

A total of 160 statistical models were generated by adding different error sets. For each of these error sets the σ was between 0.06 to 7.06. Absolutely these160 new models were deviating more or less from the nature model and from each other due to variation of the σ and the distribution of the errors within the error sets.
At this point we have to note again, that the error sets used have normal distribution with average 0 and standard deviation 1.[2,3,4]
Here we may list the expected relations:

- We expect differences between nature model and the statistical obtained model.
- The model error is assumed to increase with increasing standard deviation of the error sets.
- Further on we expect variations of the model error depending on the distribution of the values within the error set and so affecting different runs within the statistical DoE sample.

Similar result can be observed when sorting the runs within the experiment or when reducing the number of runs to half factorial design or replication are removed from statistical DoE sample. These cases are not in focus of our investigation.

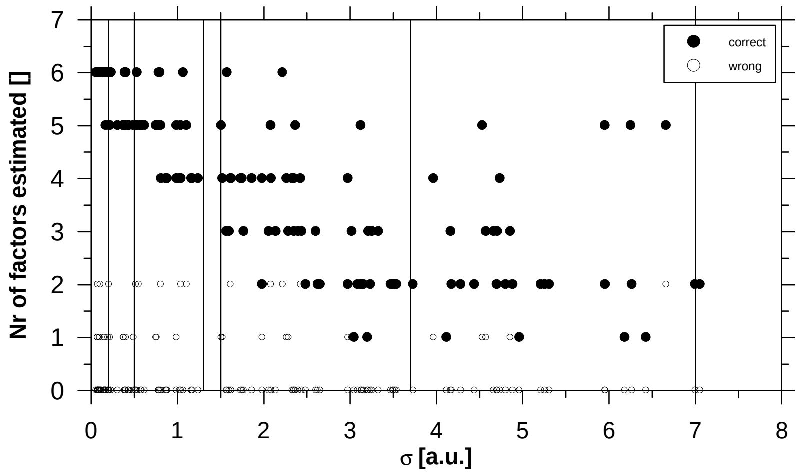


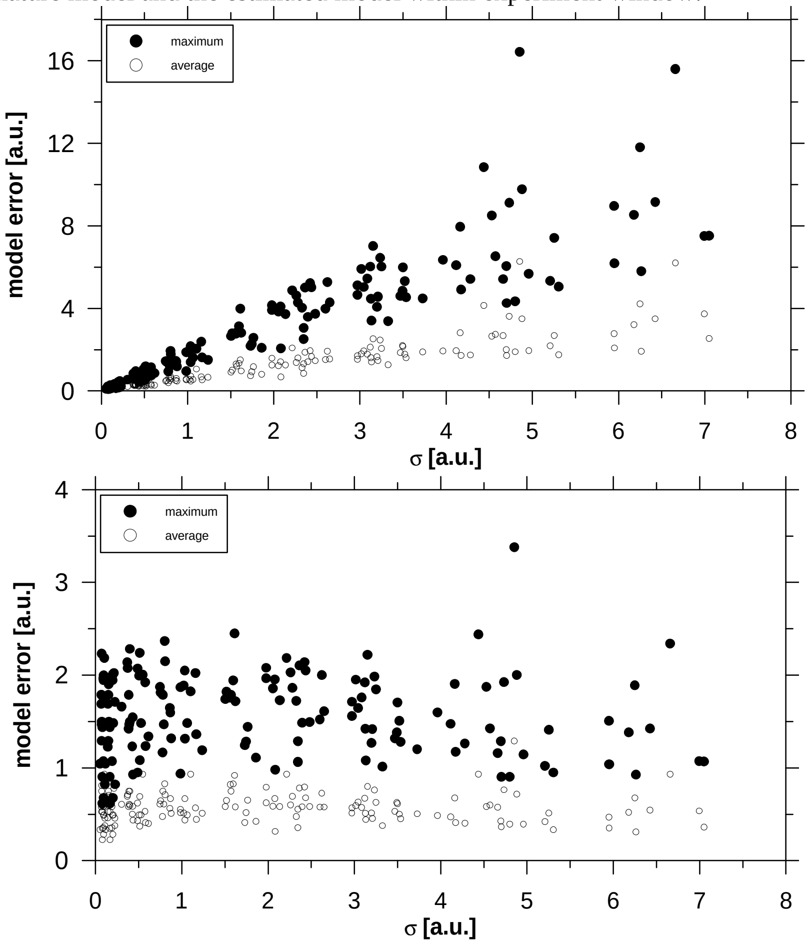
Fig.3 - Nr of factors in the model as function of standard deviation σ of the applied error set. The vertical lines indicate the effect of particular factor as listed in Table 1.

As shown in Fig.3 the sample standard deviation of the applied error sets was varied from approx. 0.06 up to 7.06 and the number of factors identified by the ANOVA analysis with application of the rules mentioned before was estimated. You can clearly see, that up to σ =0.81 all 6 significant factors are found, but in some cases also one “ghost factor” which does not belong to the model is found.

Factor	Effect
A	-3.7
B	1.5
C	0.5
D	7.0
AD	-1.3
CD	0.2

Table 1 - factors involved in the "nature model" and their effects.

The number of effects involved in the model decreases with increasing σ. Roughly the factor vanishes from the model when the σ is 3-4 times the effect of this factor - more precisely the factor vanishes when the effect of the factor cannot be estimated as significant at the probability level α. Roughly we may describe the situation as error distribution affects the center points in a way that the standard deviation of the model increases above the effect of factor.
The quality of final process model is definitely affected by any missing factor, but their contribution to the model are not identical, so the number of factors is less important and the most important criteria for the model quality is the difference between the statistical model and nature model. We decided to quantify the model error as the biggest difference between the nature model and the estimated model within experiment window.



a,

b,

Fig.4 – **a**, Error of estimated model as function of the standard deviation σ **b**, Normalized errors of estimated model.

As expected the error of the model increases with increasing σ of the error function. For our conclusion the more important aspect is the ratio between the model error and the σ itself as shown in Fig. 4b.
As one can see, the average over the 20 error sets is independent from σ and the maximum can reach a ratio of up to 3.45.

Here we want to emphasize the consequence of this result: The error of the process model is up to 3.45 times higher then the standard deviation of the measurement method used for the experiment, if the variation of the process itself is omitted.

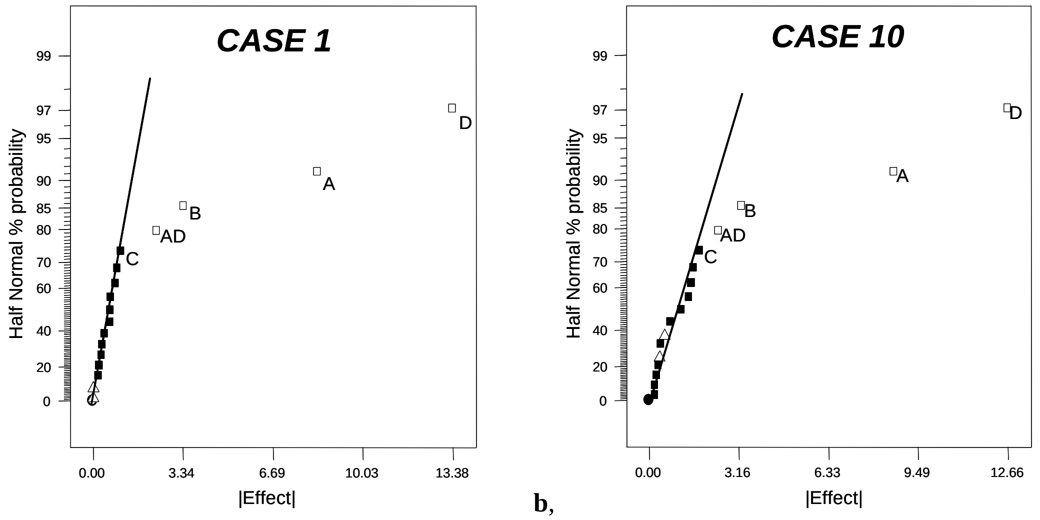
Fortunately this disadvantage can be reduced by replication, averaging of several measurements and last but not least increasing of the effect by change of the experiment process window.

EFFECT OF CENTER POINT ERROR

As one can see in Fig.3 and 4, the model error is introduced by the error set, but not only the sample standard deviation of the error set plays a role. Obviously the distribution of the error within the error set is also an important factor influencing the error of the model. The first term we want to investigate, is the impact on the model, when changing the standard deviation only for the 3 center points. For the estimation of the effect we use one error set containing 19 normally distributed numbers with σ=0.993. As reference, the error set is applied to the nature model as obtained from the random number generator and the resulting data are evaluated in order to obtain the error of the model. Subsequently the set is ordered and three values from the center of the sorted data are attached to the center points. To avoid a change of the standard deviation of the whole error set, the values are swapped with the values used for the center points in first case (see Table 2). Errors sets prepared this way were applied to the nature model and the resulting error was estimated for the case 2. Subsequently for the each next evaluation the center points were swapped with the next 3 points with slightly higher σ and the change in the model error was observed as function of the σ estimated at center points but keeping the total σ constant at 0.993. There are 969 combinations for the 3 center point within the error set of 19 values. Adding the condition of constant curvature, the number of combinations reduces dramatically. In order to keep the curvature effect around zero and thus not significant, we take from the sorted error set the items 9,10 and 11 as center points for the error set with the lowest σ (CP). Second lowest is the combination 8, 10, 12, third lowest 7, 10, 13 and so on. This approach provides us 10 DoE data sets for ANOVA evaluation. First three of them can you see in Table 2.

Run Nr.	Original	Sorted	Case 1	Case2
1	-0.576	-1.969	-0.576	-0.576
2	-0.292	-1.900	-0.292	-1.263
3	1.168	-1.263	1.168	1.168
4	-0.070	-0.917	-0.879	-0.879
5	1.183	-0.879	1.183	1.183
6	0.370	-0.576	0.370	0.370
7	1.381	-0.386	1.381	1.381
8	0.141	-0.292	0.564	0.141
9 – CP1	-1.263	-0.084	**-0.084**	**-0.292**
10 – CP2	-0.879	-0.070	**-0.070**	**-0.070**
11 – CP3	0.564	0.141	**0.141**	**0.311**
12	0.311	0.311	0.311	0.564
13	-1.969	0.318	-1.969	-1.969
14	-0.084	0.370	-1.263	-0.084
15	1.137	0.564	1.137	1.137
16	0.318	1.137	0.318	0.318
17	-0.386	1.168	-0.386	-0.386
18	-0.917	1.183	-0.917	-0.917
19	-1.900	1.381	-1.900	-1.900
s **(CP)**	***0.964***	***0.126***	***0.126***	***0.305***

Table 2 – Experiment set up - variation of the σ of center points at constant error set σ =0.993. The case 1 shows the effect of smallest possible σ estimated at center points. Here the center point at run 9 is swapped with the run 14 in original data, run 10 swapped with run 4 and run 11 swapped with run 8.



a, **b**,

Fig. 5 – Half normal plot illustrating the effect of the σ ?variation at center points at constant error set σ=0.993. a, for center point σ=0.126 b, for center point σ=1.681. Notice please the slight shift in the position of the center point symbols represented by open triangle and at the same time the remaining significant factors D, A, B, AD.

ANOVA Analysis of the DoE sets shows us, how much the results of a DoE analysis depend on the standard deviation estimated at the center points.
As shown in Fig. 5 the variation of the center point standard deviation does not affect the model significantly. The number of significant factors remains the same over broad range of σ(CP) (cf Fig.7). The reason of the invariance of the model against the center point variation is the relatively huge number of effects not involved in the model and their contribution to the mean error. In the Fig. 5a and 5b the derivative of the fit line drawn in the viewgraph represents an average standard deviation over the model. The difference in the derivative between the Fig. 5a and 5b indicates slight differences in the standard deviation of the sample standard deviation.

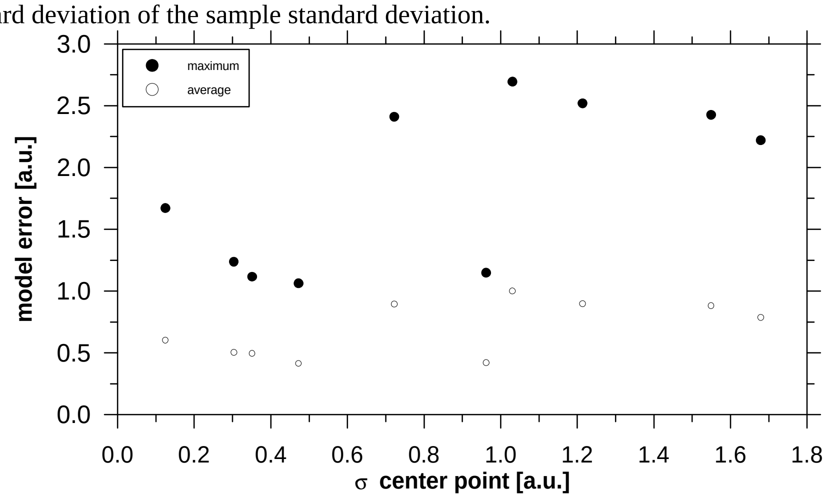


Fig.6 – maximum and average model error as function of the center point σ at constant error set σ=0.993.

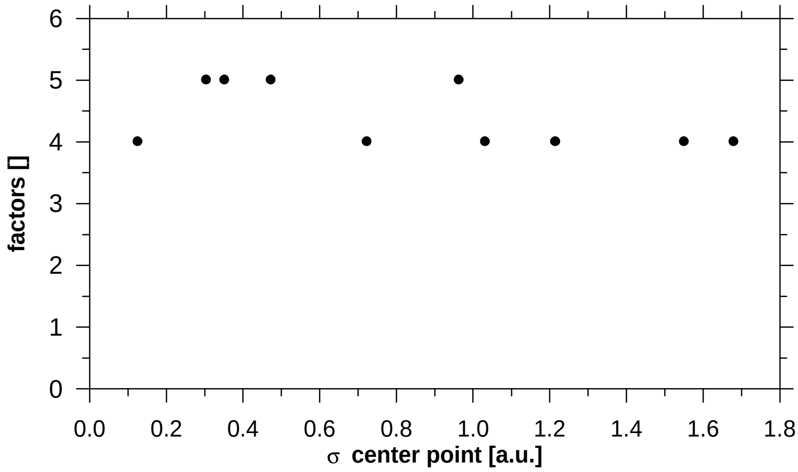


Fig.7 – number of factors as function of the center point σ.

The correlation between Fig. 6 and Fig 7 shows, that there is a non negligible contribution of the 5th factor **C** influencing the maximum error of the statistical model. However this factor is approaching the α=0.05 limit, so in several cases the factor is not significant. Since the maximum model error is about 2.7 times the standard deviation calculated from the whole error set, there has to be an additional contribution to the model error besides the center points only. Additional model error is in our opinion caused by:

- Variation of the distribution for constant s
- Permutation of the sequence within the error set
- Violation of normality within the error set

The 2nd item will be discussed in the next section; the other two won't be discussed in this paper.

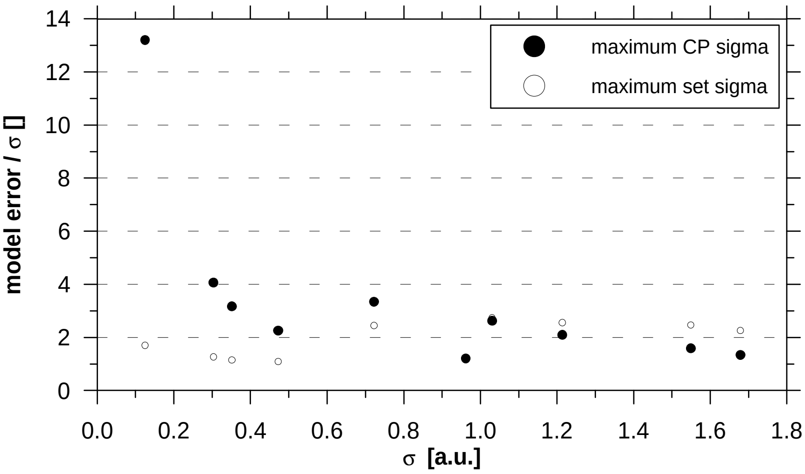


Fig. 8 – Comparison of the relative model error calculated using the error set σ and the center point σ.

EFFECT OF RANDOMIZATION/SORTING

There are 19! combinations of the error set, from which $16!\cdot\binom{19}{16}$ are causing mainly variations of the factors A, B, C, D in our model and only $\binom{19}{3}$ are influencing the center points. Since the number of DoE experiment needed in order to evaluate the whole effect is too high, we will investigate the effect of sorting with respect to few experiments only. With respect to the number of possible permutations, we will investigate only very limited group of them and focus on the link between the change in the error set and the effect caused. This evaluation provides us enough input about the possible effect of the randomization/error set sorting on the whole model.

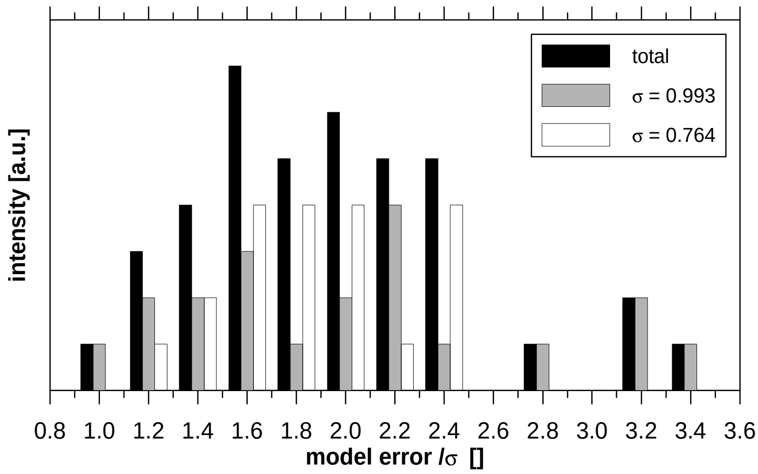
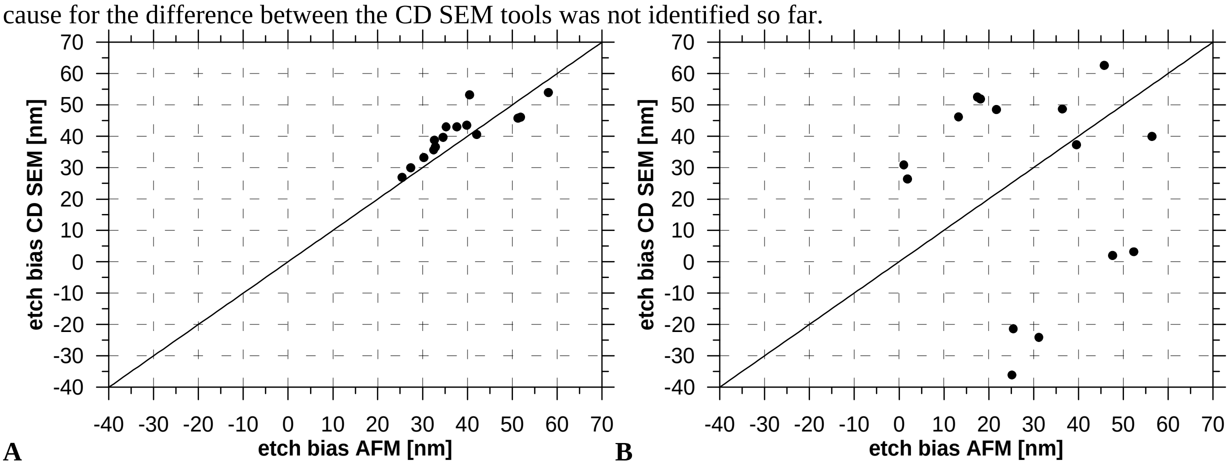


Fig. 9 – Distribution of the relative model error for various permutation of the error set. In order to check the influence of error set permutation two error sets with σ = 0.763 and σ = 0,993 were used. Within the first error set only the vertices were permuted (all points except center points), in the second one the whole error set was permuted (all 19 data points). The distribution of both data sets marked as "total" is shown as well, in the viewgraph.

The permutation of the error set is significantly contributing to the total model error. The variation of the model error due to error set permutation ranges between 1 – 3.5 σ. Since the total number of permutation is very huge ($19! = 1.21\ 10^{17}$), we were able to check only very small sample which definitely does not cover the whole range. However the number of experiments done was big enough for rough estimation of the contribution of the vertices permutation to the model error for above mentioned DoE model.

CD SEM REPEATABILITY/ PRECISION

The motivation for the investigation discussed in the previous chapters was a discrepancy observed for mean CD and CD uniformity between two different CD SEM tools with different performance and time stability. Also we reported in our previous paper[6], that etch bias estimation using CD SEM was found to be non reproducible and did not show good correlation to the etch bias measured by AFM. The resist measurement was identified as the root cause for the scattering, since the correlations of post Cr etch measurement data was at an acceptable level.
Thus the most important question was, whether both CD SEM tools are not corresponding to the atomic force microscopy (AFM) measurement. Figure 10 shows the correlation between the CD SEM and AFM tool for both CDSEM tools. The measurement done using the tool **A** correlates well to the AFM measurement obviously. The root cause for the difference between the CD SEM tools was not identified so far.



A **B**

Fig.10 – correlation between the etch bias estimated by mean of CD SEM and AFM tools. The obvious difference between both viewgraphs indicates that the resist CD measurement provides different values, which correlate to the AFM measured values in case of tool **A**, but does not in the case of tool **B**.

Another very important aspect for etch process optimization is shown in Fig 11. The goal of the process development is to reduce the systematic signatures in CD uniformity. Unfortunately any measurement error ε_m directly affects the CDU as well as the CDU signature (“footprint”). At a certain level of ε_m the footprint can not be clearly identified and so any further process optimization gets very complicated if not impossible.

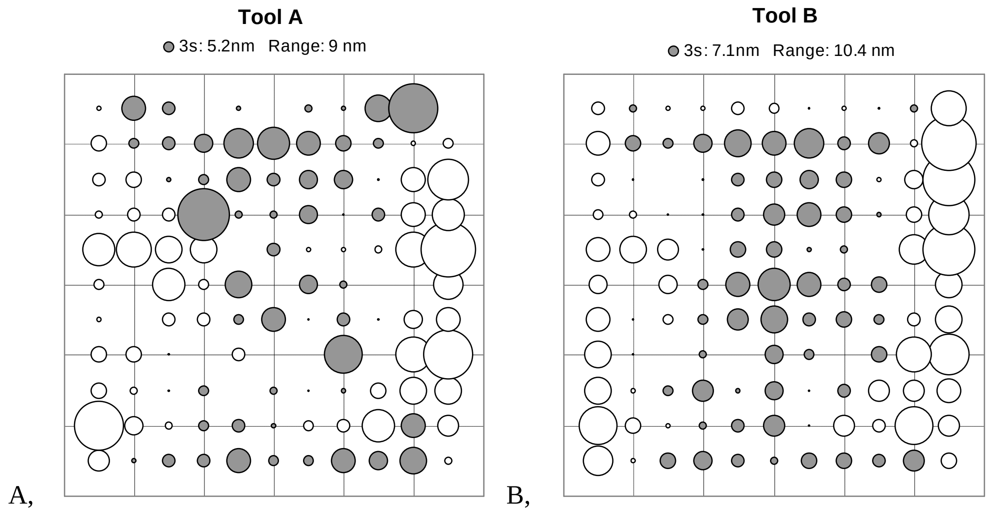


Fig.11- comparison of CD Uniformity measurement using different CD SEM tools (here marked as tool A – tool B); The error contribution is significantly lower in case of tool A which leads to lower CD Uniformity as well.

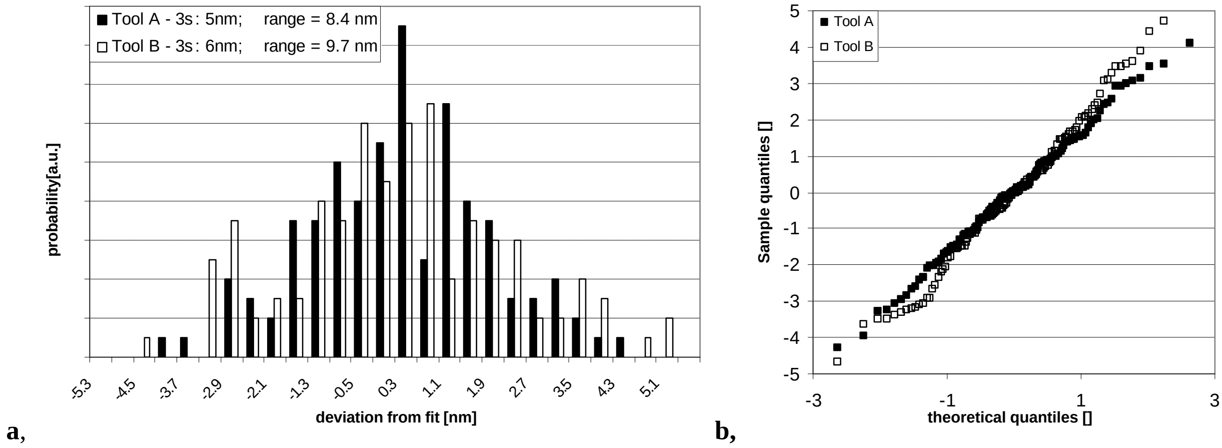


Fig.12 - **a**, Comparison of the residuals distribution measured on the same mask using tool A and tool B. The residuals are estimated as the difference between the measured CD value and the CD fit. The residuals of the measurement are assumed to be normally distributed, if the model fits well the footprint of the mask, since the error of the measurement tool is normally distributed. **b**, Normality check of the residuals error by Q-Q plot

IN order to estimate the difference in $e_\mathbf{m}$ between tools **A** and **B**, CDU measurement on one mask were compared as shown in Fig.11. By fitting the CDU footprint the residuals were estimated as difference between the fit and measured value for each point and the distribution of the residuals was calculated (Fig.12). The residuals are not only caused by ε_m, but also by local CD variation as well as line edge roughness, since the measurement position can slightly drift. The experimental data tell us, that the same process observed using the measurement from tool **A** shows in average 2 additional effects and the statistical model is cleaner then the one for tool **B**. In addition a statistical model is found for 3 responses, where the tool **B** does not provide any reasonable model.

CONCLUSIONS

In this paper we derived a formalism to understand the influence of measurement errors in statistical models for etch development. As an example we discussed the analysis of the 4 factor full factorial design with 3 center points. Our most important observation was that in our Monte Carlo simulations the model error scales linearly with the measurement error. The ratio between the precision of the measurement and the model error was estimated and we found the maximum relative model error to be about 3.5 times the sample standard deviation of the error set. This result is based on the assumption of no systematic error in the measurement, no variation of the process and ideal long term stability of the tools. We suppose when these factors are included the model error will be even larger.

It is interesting to note that the factor of 3.5 is in the range which is usually associated with measurement tool capability. This means that the measurement tool precision indeed has to meet the internal requirements of the unit process. Thus we have to emphasize that the process development places much higher demands on metrology tools than the ITRS roadmap and the capability to measure the final product satisfactory is not the full story.

Although both investigated tools fulfill the requirements of the ITRS roadmap their performance for process development was found to differ significantly. We showed this effect by analyzing the process/metrology errors on the model error in different cases. As a result we found that from the X effects detected by one tool only X-2 could be detected by another tool only due to the better short term repeatability of the first one. In some cases it was impossible to find a model for tool B while there was one identified with tool A.

Further investigations will focus on enlarging the number of Monte Carlo runs to widen the statistical data base with the final goal to quantify the effect of the measurement errors on process parameters like feature uniformity or expected stability of the found process

REFERENCES

1. D.C. Montgomery: Design and Analysis of Experiments; John Wiley; 2001; 5th ed.; ISBN 0-471-31649-0
2. M.Matsumoto, T. Nishimura: ACM Transactions on Modeling and Computer Simulation (TOMACS); Vol. 8; Issue 1; (1998); p.3
3. W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling: Numerical Recipes in Pascal; Cambridge University Press; 1989; ISBN 0-521-37516-9
4. D.E. Knuth: The Art of Computer Programming: Vol.2; Addison -Wesley; 1998; 3rd ed.; ISBN 0-201-89684-2
5. M.J. Wichura: Algorithm AS 241: The Percentage Points of the Normal Distribution; Applied Statistics; Vol. 37; 1988; p. 477-484
6. P. Nesladek, N. Falk, A. Wiswesser, R. Koch, B. Sass: *25th Annual BACUS Symposium on Photomask Technology*, Proc. SPIE Vol. 5992; pp. 190; (2005)