

# The Flash Memory battle: How low can we go?

<sup>a</sup>Eelco van Setten, <sup>a</sup>Onno Wismans, <sup>a</sup>Kees Grim, <sup>a</sup>Jo Finders, <sup>b</sup>Mircea Dusa,

<sup>c</sup>Robert Birkner, <sup>c</sup>Rigo Richter, <sup>c</sup>Thomas Scherübl

<sup>a</sup>ASML Netherlands B.V., De Run 6501, 5504 DR Veldhoven, The Netherlands

<sup>b</sup>ASML TDC, 4800 Great America Parkway – Suite 500, Santa Clara, CA 95054, USA

<sup>c</sup>Carl Zeiss SMS GmbH, 07740 Jena, Germany

## ABSTRACT

With the introduction of the XT:1900Gi the limit of the water based hyper-NA immersion lithography has been reached in terms of resolution. With a numerical aperture of 1.35 a single expose resolution of 36.5nm half pitch has been demonstrated. However the practical resolution limit in production will be closer to 40nm half pitch, without having to go to double patterning like strategies. In the relentless Flash memory market the performance of the exposure tool is stretched to the limit for a competitive advantage and cost-effective product.

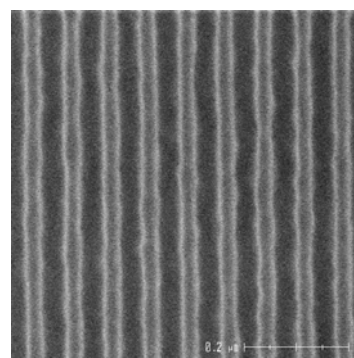
In this paper we will present the results of an experimental study of the resolution limit of the NAND-Flash Memory Gate layer for a production-worthy process on the XT:1900Gi. The entire gate layer will be qualified in terms of full wafer CD uniformity, aberration sensitivities for the different wordlines and feature-center placement errors for 38, 39, 40 and 43nm half pitch design rule. In this study we will also compare the performance of a binary intensity mask to a 6% attenuated phase shift mask and look at strategies to maximize Depth of Focus, and to de-sensitize the gate layer for lens aberrations and placement errors. The mask is one of the dominant contributors to the CD uniformity budget of the flash gate layer. Therefore the wafer measurements are compared to aerial image measurements of the mask using AIMS<sup>TM</sup> 45-193i to separate the mask contribution from the scanner contribution to the final imaging performance.

Keywords: NAND-Flash, hyper-NA, immersion lithography, AIMS

## 1. Introduction

Flash memory has become one of the most important segments of the semiconductor industry in recent years. Flash memory is also an important driver of the lithography roadmap, with its dramatic acceleration in dimensional shrink, pushing for ever smaller feature sizes. With the introduction of the XT:1900Gi the limit of water based hyper-NA immersion lithography has been reached in terms of resolution. With a numerical aperture of 1.35 a single expose resolution of 36.5nm half pitch has been demonstrated<sup>1</sup> (see Figure 1). However, the practical resolution limit for high volume NAND-Flash manufacturing will be in the 38 to 40nm half pitch range for a single expose process.

Figure 1 36.5nm half pitch dense L/S ( $k_1 = 0.255$ ) resolved on an XT:1900Gi immersion scanner: Very close to the theoretical resolution limit of 35.8nm HP for water based hyper-NA immersion lithography



The NAND-type flash memory cell (Figure 2) is organized in separate modules (pages) containing 32 wordline transistors (wordline poly gates) with 2 select transistors (select poly gates) and with one source and drain contact for the entire 32 page module. The array-core zone consisting of 32 equal line-spaces allows for aggressive low- $k_1$  imaging with extreme off-axis illumination. However, the presence of the two mirrored SG lines and the SG-SG gap, break up

the array pitch regularity and create a discontinuity in layout topology. Dusa et al<sup>2</sup> showed that this discontinuity introduces an asymmetry in the near field diffracted intensity under the left and right edges of the SG line, as well as under the adjacent next 4 to 5 wordlines (see Figure 3). The lithographic performance of the mask layout is affected by this cross-coupling effect, which induces feature-center placement errors through dose and focus. This feature-center placement error, or intra-module overlay error, consumes a significant part of the total overlay budget. This means that the single expose scaling limit for NAND-Flash is to a large extent determined by the resolution limit and CD control of the array-core zone, and the placement error of the select gates and first few wordlines. Lens heating effects should, of course, also be well controlled, but this falls outside the scope of this paper.

In this paper we investigated the practical resolution limit for NAND-Flash devices by looking at the CD control and overlay budget of the gate layer. To study these effects at mask level, and to discriminate between the mask and the scanner, a test mask has been measured using AIMS<sup>TM</sup> 45-193i<sup>3</sup>. Exposures were performed with an XT:1900Gi hyper-NA exposure tool. The test mask contains NAND-type flash memory clips, at multiple features sizes patterned in MoSi, and MoSi covered with a chrome layer.

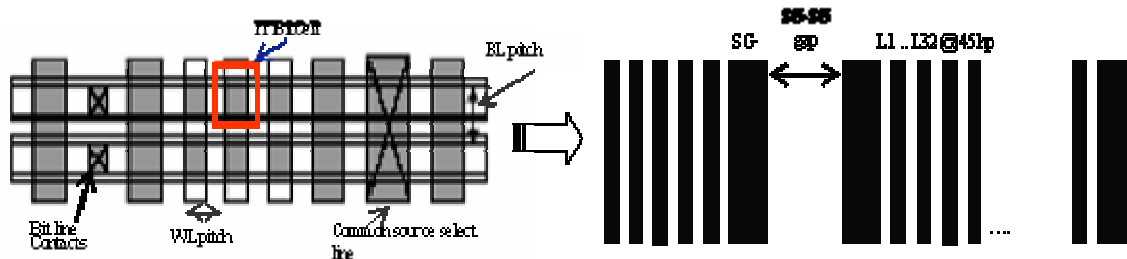


Figure 2 Left: Flash memory core layout for critical layers: poly wordline and interconnect bitline and contacts. Layout simplicity with unidirectional regularity is noticeable. Right: Characteristic topology of poly wordline layout with line/space regularity broken by SG and SG-SG gap. Design rule defines SG and SG-SG gap dimensions as multiple integers of half pitch.

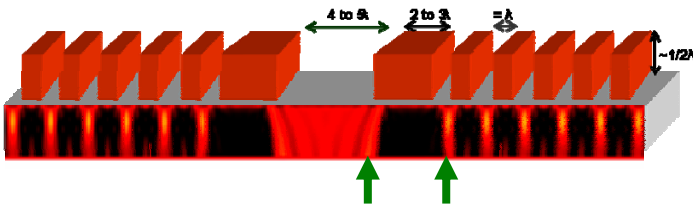


Figure 3 Scattered near field diffracted intensity for TE-polarized illumination from flash wordline mask pattern. An asymmetry under the left and right edges of the SG line and adjacent wordlines is visible (e.g. compare the intensity at the green arrows).

## 2. Experimental conditions

All exposures and measurements were done with a test mask from AMTC -Toppan. The film stack of the mask consists of a 680 Å MoSi layer covered by a 580 Å chrome layer. The MoSi-features (after patterning) have an average phase shift of 180° and transmission of 6.15%. The binary structures are patterned in the thick MoSi + Cr stack as opposed to traditional Chrome-On-Glass (COG) binary masks, which do not have a MoSi layer between the Cr top and the glass substrate. In this paper, the term ‘Binary’ will be used for the ternary stack as described above, while ‘Attenuated’ refers to the 6.15% transmission of the MoSi-features. A schematic representation of the mask stack on our test reticle can be found in Figure 4 below.

The mask has been measured extensively with an AIMS<sup>TM</sup> 45-193i aerial image measurement system from Zeiss. The AIMS<sup>TM</sup> 45-193i has a maximum scanner equivalent NA of 1.4, which enables the emulation of the latest generation of hyper-NA immersion scanners. Furthermore it is possible to use polarization and multiple off-axis illumination modes, equivalent to scanner illumination modes. The aerial image measurements were done at NA = 1.35 and  $\sigma = 0.78 / 0.98$  Dipole X illumination with 35° opening angle and linearly polarized light (Y-polarized) in ‘Scanner Mode’<sup>4</sup> to capture the vector effects associated with hyper-NA imaging. In the “Scanner mode” the aerial image in resist will be generated

as a result of a combination of actual image measurements in the scalar mode (or “AIMS™ mode) and a Zeiss proprietary algorithm. Additionally, the user has to input the refractive index of the resist. In this work the refractive index was set to 1.7, which is equivalent to the refractive index of the photo resist used during the exposures.

The exposures were performed with an ASML XT:1900Gi hyper-NA exposure tool with a maximum NA of 1.35. The illumination conditions were optimized for 38nm, 40nm and 43nm half pitch using NA = 1.35 and dipole X illumination with 35° opening angle and linearly polarized light (Y-polarized), see Table 1. The exposures were done on bare Si wafers coated with 93nm Brewer Science ARC93SR BARC, 95nm TOK TARFPi6001 resist and a 90nm JSR TCX041 top coat. All wafer measurements were carried out by CD SEM (Hitachi CG-4000). The difference in sigma settings between the AIMS™ measurements and the exposures has been chosen on purpose to closely match the intensity profile of the scanner illumination source. The AIMS™ source can be regarded as top-hat illumination, whereas the scanner source has a more Gaussian intensity distribution. Furthermore, sigma-inner/ -outer on the scanner are defined as the radius where 10% / 90% of the total energy is encircled, whereas the top-hat sigma settings are defined at the edges (0% / 100% of the encircled energy).

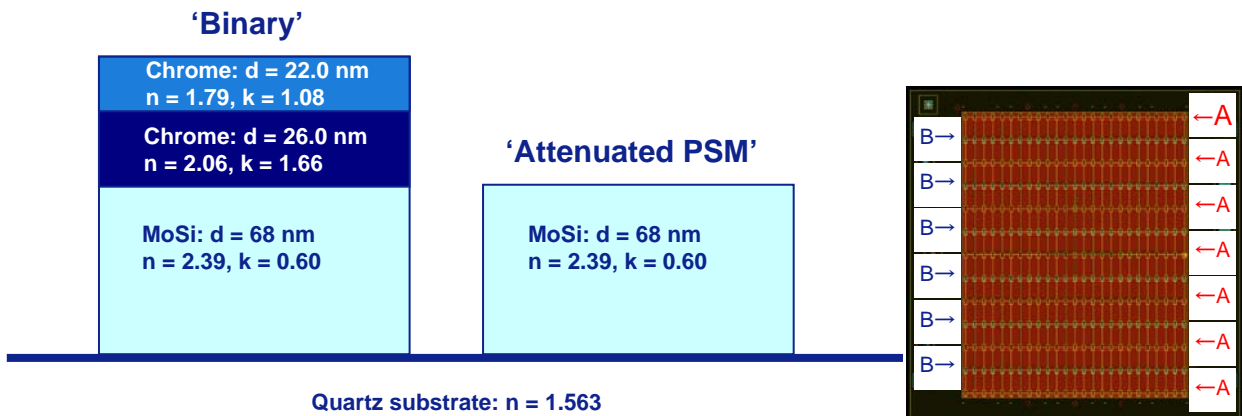


Figure 4 Left: Schematic representation of the mask stacks and their optical constants used in the experiments. Right: Test mask with alternating rows of attenuated (A) and binary (B) features.

	Scanner				AIMS				Mask OPC	
	NA	Sigma	Source	Polarization	NA	Sigma	Source	Polarization	BIN	ATT
38 nm DL	1.35	0.885/0.985	Dipole-X-35	Y-pol	-	-	-	-	NO	NO
39 nm DL	1.35	0.82/0.97	Dipole-X-35	Y-pol	1.35	0.78/0.98	Dipole-X-35	Y-pol	YES	YES
40 nm DL	1.35	0.82/0.97	Dipole-X-35	Y-pol	1.35	0.78/0.98	Dipole-X-35	Y-pol	YES	YES
43 nm DL	1.35	0.75/0.91	Dipole-X-35	Y-pol	-	-	-	-	NO	NO

Table 1 Overview of illumination settings per feature size used for the exposures and AIMS™ measurements. The column ‘Mask OPC’ indicates if (basic) OPC was applied for this particular illumination condition.

### 3. NAND-type flash memory cell

In the next section the measurement results from NAND-type flash memory structures are presented.

#### 3.1. Flash wordline mask pattern

Figure 5 shows the features from the flash wordline mask pattern that were evaluated in this study and their design CD’s per CD node: The central space (SGSG), select gates (SG), wordline 1, 2 and 7, and the spaces between SG and WL1 (SP0) and between WL1 and WL2 (SP1). All 7 features were measured with AIMS™ on mask level and with CD-SEM on wafer level in resist on both sides of the central space (only the right side is shown). Wordline 7 is regarded as fully dense and is used as reference to print the gate layer on target. A basic OPC treatment has been applied using the thin mask approximation (Kirchhoff) and a lumped parameter resist model for the settings indicated in Table 1.

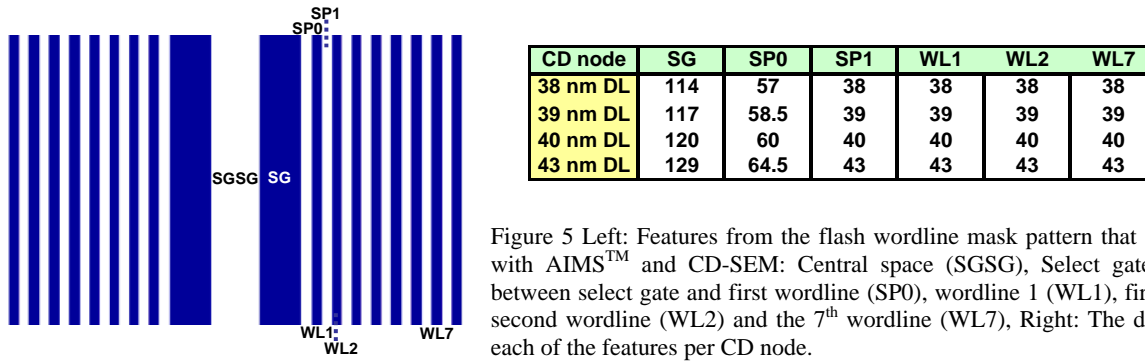


Figure 5 Left: Features from the flash wordline mask pattern that were evaluated with AIMS<sup>TM</sup> and CD-SEM: Central space (SGSG), Select gates (SG), space between select gate and first wordline (SP0), wordline 1 (WL1), first space (SP1), second wordline (WL2) and the 7<sup>th</sup> wordline (WL7), Right: The design CD's for each of the features per CD node.

The test mask contains per feature size two designs with a different design rule for SP0, one of the most critical features in the gate layer. In the first design SP0 has a target CD which is equal to 1.5 times the design rule half pitch (e.g. 1.5 x 40nm = 60nm space). In the second design the space has been increased to 2.0 times the design rule half pitch (see Figure 6). In both cases OPC has been applied to print all other features to size, which is the design rule half pitch for all wordlines and spaces, except SP0, for the select gates 3 times the half pitch and for the central space 6 times the half pitch.

Figure 6 also shows the aerial image contour of the flash wordline mask pattern that was measured with AIMS<sup>TM</sup> 45-193i on top of a CD-SEM image of the identical module. This picture exemplifies that the AIMS<sup>TM</sup> 45-193i is capable of accurately capturing the optical effects of a complex 2D-structure, closely matching the final image in resist.

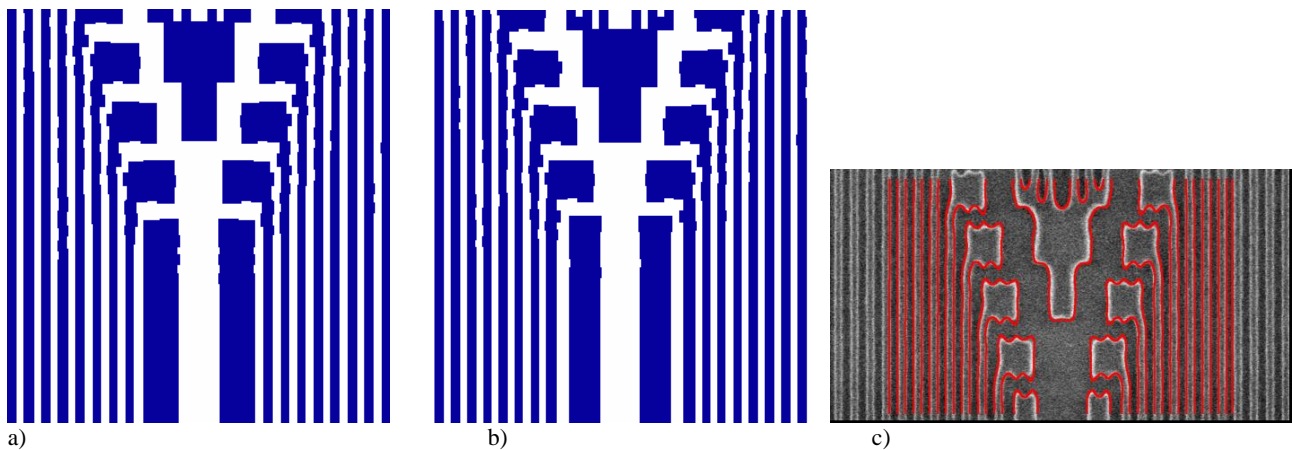


Figure 6 Different designs for SP0: a) 1.5 x minimum pitch, b) 2.0 x minimum pitch, c) Aerial image contour of SP0 1.5 design captured with AIMS<sup>TM</sup> 45-193i (red lines) on top of CD-SEM image. The landing pads and part of a dummy structure are also shown in these pictures

### 3.2. Mask CDU measured with AIMS<sup>TM</sup>

The qualification of a mask is usually done using CD-SEM measurements. This gives a good indication of the quality of the mask in terms of linewidth variations and mean-to-nominal CD. However, additional contributors to the CD on the wafer, such as side wall angle variations, transmission variations and phase shift errors remain invisible and have to be qualified with other techniques. With AIMS<sup>TM</sup> it is possible to measure the CD variation on mask level for a specific illumination mode, including all variations and material imperfections that are not picked up by CD-SEM<sup>5</sup>. Furthermore, the AIMS<sup>TM</sup> measurement includes the MEEF, which is dependent on the illumination setting and feature size. The effect of resist contrast, resulting in an increased MEEF on wafer level, can not be included, although this may become possible in future software releases.

Figure 7 shows across field CD fingerprints for the binary 40nm half pitch design rule (SP0 1.5 design) flash wordline pattern measured with CD-SEM in resist on the wafer and with AIMS<sup>TM</sup> on mask level. To determine the aerial image CD's of the features in the gate layer with AIMS<sup>TM</sup> first wordline 7 is evaluated. A threshold value is determined to print WL7 to size (i.e. 40nm) for the center position on the mask. This threshold is used to calculate the CD's of the other features of interest, both within the wordline pattern and across the mask. For all evaluated features the fingerprints match well, which indicates that the CD variation on the mask dominates the CD uniformity on the wafer. This not only holds for the dense wordlines and spaces like WL7, WL2 and SP1, but also for semi-dense or semi-isolated features like WL1, SP0 and the select gates, which are much more sensitive for focus errors on the wafer. Figure 8 shows the mean CD and the across field CD uniformity corresponding to the evaluated features in Figure 7 for both the AIMS<sup>TM</sup> and CD-SEM measurements. The feature sizes in the gate layer measured by CD-SEM and AIMS<sup>TM</sup> correspond well to each other. The maximum deviation is found to be 4.4nm, which is surprisingly small taking into account that no effort has been made to calibrate the metrology tools, the CD-SEM measurements are done in resist whereas the AIMS<sup>TM</sup> measurements are done in air, and that the measurements were not performed on exactly the same location in the flash structure (same feature but not the exact same (x,y)-coordinate). The across field CD uniformity of the wordline pattern on the wafer is larger than the mask CDU measured with AIMS<sup>TM</sup>. This is as expected since the wafer measurements include across field dose and focus errors from the scanner and a higher MEEF due to the photoresist layer. However, the graph clearly shows that the mask CDU is the dominant contributor to the final CDU on the wafer. CDU differences between the features in resist closely follow the CDU differences on mask level.

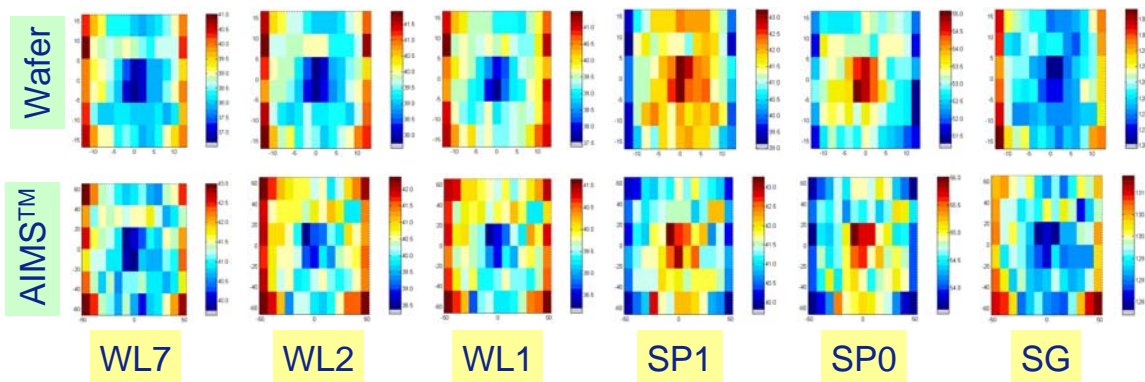


Figure 7 Across field CD fingerprints for 40nm half pitch flash wordline pattern measured on the wafer with CD-SEM and on mask level with AIMS<sup>TM</sup>.

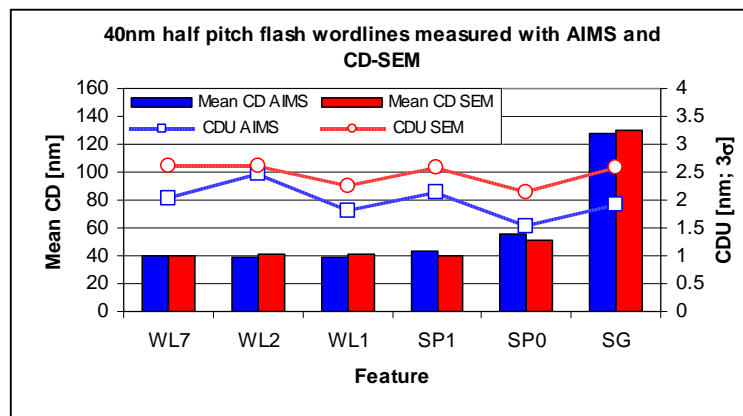


Figure 8 Mean CD (left y-axis) and across field CD uniformity (right y-axis) of binary 40nm half pitch flash wordlines measured with AIMS<sup>TM</sup> on mask level (numbers are at 1x) and with CD-SEM in resist on the wafer.

### 3.3. MEEF, process windows and CD uniformity predictions

As shown above, the mask CD uniformity is a dominant contributor to the CDU on the wafer. Therefore the mask error enhancement factor (MEEF) is an important parameter, which should be as low as possible in order to meet the CD uniformity requirements. The MEEF has been measured for 38, 39, 40 and 43nm dense L/S gratings, which were exposed using the illumination settings in Table 1.

Table 2 shows the MEEF for both the attenuated and binary modules, and the required mask CD uniformity (at 4x, 3σ) if we assume that 5% of the CD node can be consumed by reticle CD variations. The MEEF of the 38nm dense L/S is kept under control by using extremely high sigma settings and a small pole width. The table also shows that the binary features have a slightly lower MEEF than the attenuated features.

MEEF	MEASURED (SEM)		MASK CDU (4X)	
	Att	Bin	Att	Bin
38 nm DL	3.00	2.80	2.53	2.71
39 nm DL	2.80	2.80	2.79	2.79
40 nm DL	2.50	2.40	3.20	3.33
43 nm DL	2.30	2.10	3.74	4.10

Table 2 Measured MEEF for 38, 39, 40 and 43nm dense L/S using dipole illumination and the required mask CDU (at 4x, 3σ) corresponding to 5% of the CD node. NOTE, that the sigma settings have been adjusted for the 38 and 43nm features.

Apart from the MEEF, other basic imaging characteristics like Exposure Latitude (EL) and Depth of Focus (DoF) have been determined. Figure 9 shows the EL for all evaluated features in the flash wordline pattern for 38, 39, 40 and 43nm half pitch design rule (SP0 1.5 design). The dose-critical features (Wordlines 1, 2, 7 and SP1) all have an EL above 10% in all cases. However, it is clearly visible that the EL drops significantly when the half pitch is reduced from 43 to 40nm and from 40 to 39nm. By adjusting the sigma settings for the 38nm modules (see Table 1), a further reduction in EL can be prevented.

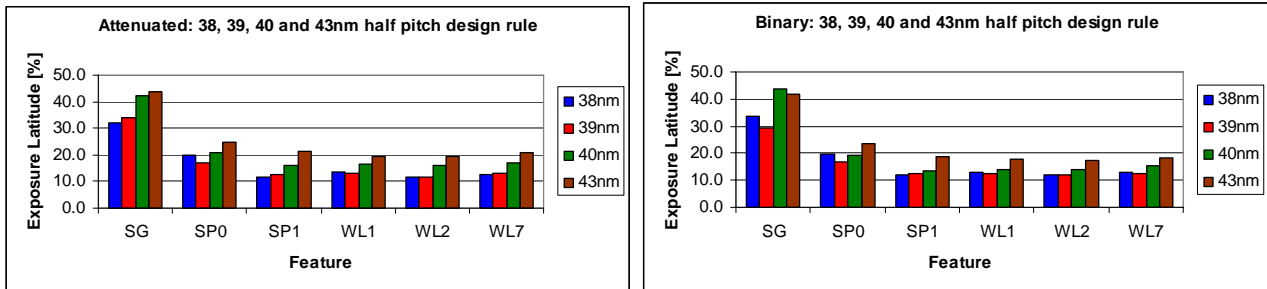


Figure 9 Measured EL for all evaluated features in the gate layer for 38, 39, 40 and 43nm half pitch design rule (SP0 1.5 design). Left: Attenuated module, Right: Binary module.

If we take the most dose-critical feature in the gate layer per CD node (which is WL2) and assume a total dose error of 0.9% 3σ we can estimate the expected contribution to CD uniformity from dose variations. This is 0.4nm 3σ for 43nm half pitch, but increases to 0.6nm 3σ for 38nm half pitch (see Table 3).

	Min. EL (WL2 )		CDU contr.	
	Att	Bin	Att	Bin
38 nm DL	11.4	11.8	0.60	0.58
39 nm DL	11.9	12.0	0.59	0.59
40 nm DL	15.9	13.5	0.44	0.52
43 nm DL	19.4	17.1	0.37	0.42

Table 3 Minimum EL in the gate layer for 38, 39, 40 and 43nm de half pitch design rule (SP0 1.5 design) and the estimated contribution to the CDU (3σ).

The DoF for all evaluated features is shown in Figure 10 below. The DoF of the 43nm module is significantly larger than the other features sizes for all evaluated features. The minimum DoF is observed for 38nm half pitch design rule and equal to 100nm. The minimum DOF of the binary modules is found to be slightly larger than the minimum DoF of the attenuated modules for all evaluated features sizes. WL1 and SP0 are in all cases DoF limiting, and thus the most focus-critical features. The focus sensitivity can be derived from the Bossung curves, which were used to calculate the DoF, by a second order polynomial fit. If we assume a total focus error of 120nm 8σ (or 45nm 3σ) and a focus offset of 5nm, we can translate the focus sensitivity into a CD error. Table 4 shows the Bossung curvature (nm/μm<sup>2</sup>) for SP0 and WL1 together with the estimated contribution to CD uniformity. SP0 is the most focus sensitive feature and, hence gives

the largest contribution to CDU. If we take the CDU relative to the targetCD than the maximum contribution is 6% for SP0 and 3% for WL1, in both cases for the 40nm attenuated module. It should be noted that the focus sensitivity of the 40nm modules is larger than the 39nm modules, although the illumination setting was optimized for 40nm half pitch.

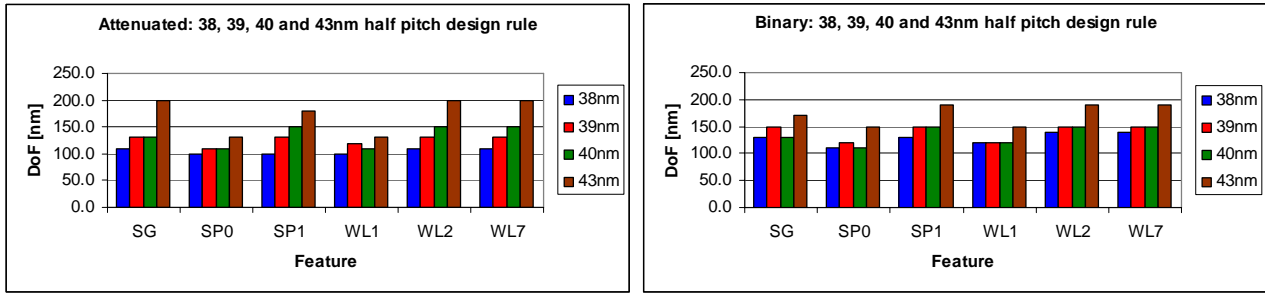


Figure 10 Measured DoF for all evaluated features in the gate layer for 38, 39, 40 and 43nm half pitch design rule (SP0 1.5 design). Left: Attenuated module, Right: Binary module.

	Bossung Curv. (SP0)		Bossung Curv. (WL1)		CDU contr. (SP0)		CDU contr. (WL1)	
	Att	Bin	Att	Bin	Att	Bin	Att	Bin
38 nm DL	-2900	-3500	2300	1800	2.19	2.65	1.74	1.36
39 nm DL	-2500	-2700	1900	1600	1.89	2.04	1.44	1.21
40 nm DL	-4600	-3400	2400	2100	3.48	2.57	1.82	1.59
43 nm DL	-1800	-2200	1500	1000	1.36	1.66	1.13	0.76

Table 4 Bossung curvature ( $\text{nm}/\mu\text{m}^2$ ) of SP0 and WL1 for 38, 39, 40 and 43nm de half pitch design rule (SP0 1.5 design) and the estimated contribution to the CDU ( $3\sigma$ ).

If we assume that the reticle, lumped focus and dose errors make up for 90% of the final CD uniformity on the wafer, we can construct a simple CDU budget and compare that to the CDU requirements. The CD uniformity requirements for the flash memory gate layer can be specified as follows. The total CD control after etch should be 10% of the node or less. If we divide this equally among litho and etch, 7.1% of the CD node (assuming Gaussian error distributions) can be allocated for the litho step. In our simple budget we take 90% of that value and arrive at a CD error budget of 6.7% of the CD node. Table 5 shows the CDU contribution from the mask (fixed at 5% of the CD node), from dose-, focus variations and the total estimated CD uniformity for the two most critical features in the gate layer: SP0 and WL1. Note that SP0 has a larger design CD (see also Figure 5) than WL1, which translates to less stringent CDU and mask requirements. If we compare this to the CDU requirements, which are specified as 6.7% of the design CD, then it becomes clear that the CDU contribution from focus and the mask are critical. It should be noted that the CDU budget ignores the influence from (odd) lens aberrations, which can lead to significant left-right CD differences across the field as will be shown later. However, in our analysis of the measured CD uniformity we found that the CD distribution of the pooled left-right populations increases only marginally with respect to the individual populations (no left-right CD offset).

CDU SP0	CDU mask		CDU dose		CDU focus		Total CDU contr.		CDU requirements	
	Att	Bin	Att	Bin	Att	Bin	Att	Bin	Att	Bin
38 nm DL	2.85	2.85	0.52	0.52	2.19	2.65	3.63	3.92	3.82	3.82
39 nm DL	2.93	2.93	0.62	0.62	1.89	2.04	3.54	3.62	3.92	3.92
40 nm DL	3.00	3.00	0.51	0.56	3.48	2.57	4.62	3.99	4.02	4.02
43 nm DL	3.23	3.23	0.47	0.49	1.36	1.66	3.53	3.66	4.32	4.32
CDU WL1	CDU mask		CDU dose		CDU focus		Total CDU contr.		CDU requirements	
	Att	Bin	Att	Bin	Att	Bin	Att	Bin	Att	Bin
38 nm DL	1.90	1.90	0.51	0.53	1.74	1.36	2.63	2.40	2.55	2.55
39 nm DL	1.95	1.95	0.54	0.57	1.44	1.21	2.48	2.37	2.61	2.61
40 nm DL	2.00	2.00	0.43	0.50	1.82	1.59	2.73	2.60	2.68	2.68
43 nm DL	2.15	2.15	0.37	0.41	1.13	0.76	2.46	2.32	2.88	2.88

Table 5 Estimated CDU contribution for 38, 39, 40 and 43nm de half pitch design rule (SP0 1.5 design) from mask, dose and focus errors and the CDU requirement specified as 6.7% of the CD node (all values are in nm  $3\sigma$ ) for SP0 and WL1. Note that the CDU requirement for SP0 is less stringent due to the larger design CD.

### 3.4. Measured CD uniformity of flash memory gate layer

The CD uniformity of the most critical features in the gate layer have been measured in detail for 40, 43 and 38nm design rule half pitch (SP0 1.5 design) binary modules, which were exposed using the illumination settings in Table 1. Reticle error correction was applied using the mask fingerprints of the 40nm binary module measured with AIMS<sup>TM</sup> (see Figure 7) for the 40nm data, but also the 43nm and 38nm wafer data. Since we did not measure the 43 or 38nm modules on the mask, we can only correct the global reticle fingerprint but not any local CD variations, reducing the effectiveness of the reticle error correction. Furthermore, the AIMS<sup>TM</sup> measurements were performed with a slightly different illumination setting and at slightly different locations in the flash module than the CD-SEM measurements. The last point holds also for the 40nm module, which means that intra-module CD variations on the mask cannot be removed effectively.

Figure 11 shows the mean CD, across wafer (AW) and across field (AF) CD uniformity of the pooled left-right population for all evaluated features in the gate layer for the 40nm and 43nm modules. The CDU target for litho of 7.1% of the design CD has also been included. The across wafer CDU has been determined on 38 fields (26.5 x 33.5mm field size) distributed over the wafer, of which 9 sites per field are measured in 36 fields and 78 sites per field in 2 fields. The fields with 78 measured sites are used to determine the across field CDU. The 40nm data shows that the AW CDU is less than 3.0nm 3 $\sigma$  for WL7, WL2, WL1 and SP1. For SP0 and SG this increases to 3-4nm 3 $\sigma$ , but these features also have a larger targetCD. The AF CDU is 3.2nm 3 $\sigma$  or below for all evaluated features and the combined AF CDU of wordlines 1, 2 and 7 on both sides of the central space is 2.5nm 3 $\sigma$ . The 43nm data shows similar results: less than 4.0nm 3 $\sigma$  AW CDU for all evaluated features and less than 3.5nm 3 $\sigma$  AF CDU in all cases. The combined AF CDU of wordlines 1, 2 and 7 on both sides of the central space is 3.7nm 3 $\sigma$ . However this number could be improved significantly by a better OPC, since there is a 2.0nm mean CD difference between WL7 and WL1 (this is 1.0nm for the 40nm module).

Figure 12 shows the AF CDU for 38nm half pitch design rule. The AF CDU is below 2.5nm 3 $\sigma$  for WL1, WL2, WL7 and SP1, although the reticle fingerprint could not be removed effectively due to a higher MEEF and the reasons mentioned above. We could verify a good image integrity of all features in the gate layer over the entire exposure field. Furthermore, the AF CDU does not change significantly when reducing the pitch from 43nm to 38nm half pitch. This indicates that the CD uniformity can be well controlled if the exposure conditions are properly optimized.

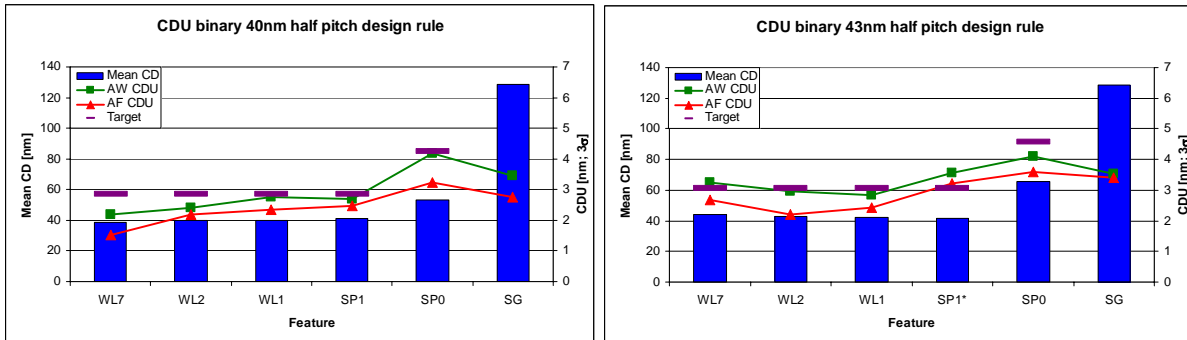


Figure 11 Mean CD, across wafer (AW) CDU, across field (AF) CDU and 7.1% CDU target for all evaluated features in the gate layer as measured with CD-SEM on the wafer, data from left and right of the central space is pooled. Left: 40nm data, Right: 43nm data, SP1: SEM data from the right side of SGSG is missing.

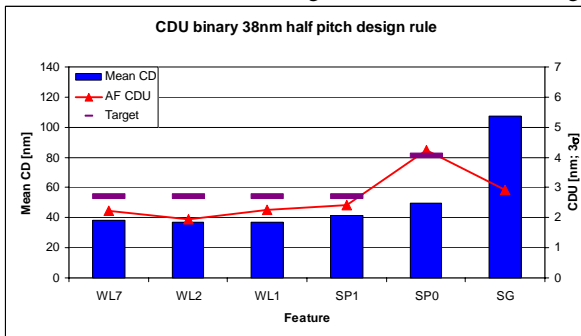


Figure 12 Mean CD, across wafer (AW) CDU, across field (AF) CDU and 7.1% CDU target for all evaluated features in the gate layer, as measured with CD-SEM on the wafer for 38nm half pitch design rule, data from left and right of the central space is pooled.

### 3.5. Feature-center placement error of flash memory gate layer

Dusa et al. predicted that the imbalance in the near-field diffraction intensity in the vicinity of the select gates would lead to a feature-center shift through focus of the select gates and the adjacent 4 to 5 wordlines. This feature-center shift can

be regarded as an intra-module overlay error coming from the design itself, in addition to intra-field and across wafer overlay errors coming from the scanner and the reticle. From the CD-SEM and AIMS<sup>TM</sup> measurements through focus the feature-center placement errors of the select gates, WL1 and WL2 have been determined. The center of the central space (denoted as SGS<sub>G</sub> in Figure 5) has been chosen as reference. A positive shift indicates that the line moves away from the center, i.e. in the positive x-direction for features on the right side of the central space and in the negative x-direction for features on the left side of the central space. Conversely, a negative shift indicates that the line moves towards the central space (see also Figure 13).

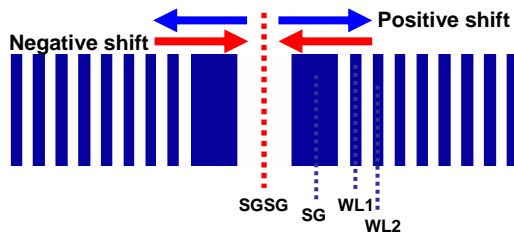


Figure 13 Definition of feature-center shift: A positive shift moves the line away from the center, a negative shift towards the center. The evaluated features are SG, WL1 and WL2.

Figure 14 shows the results for the binary modules with 40nm half pitch design rule. The select gates have a positional shift of 2nm in best focus, which increases to ~9nm when going 60nm out of focus. Wordline 1 goes in the opposite direction of the select gates and shows a shift of -7nm in best focus and ~ -11nm when going 60nm out of focus. Wordline 2 shows the smallest shift, only -4nm in best focus, which remains almost stable through focus (~ -5nm when going 60nm out of focus).

If we assume that the overlay budget is 20% of the CD node (i.e. 8nm in this case), of which a 4nm error can be tolerated for these intra-module overlay errors, then it becomes clear that the positional shift consumes almost the total overlay budget, leaving no room for additional overlay errors from the scanner and the reticle. But also if we allocate the full 8nm overlay budget to the placement errors, then the focus range should still be controlled within +/-40nm (WL1 being the limiting feature). This means that the required focus control is driven by the placement errors through focus rather than the CD variations.

Besides the lithographic implications that can be derived from Figure 14, a remarkable agreement between the AIMS<sup>TM</sup> measurements on mask level and wafer measurements with CD-SEM can be observed. The correlation plot in Figure 14 shows that the AIMS<sup>TM</sup> and CD-SEM data sets for both modules are statistically almost identical assuming a measurement error of 1.0nm 1 $\sigma$  for CD-SEM and 0.4nm 1 $\sigma$  for AIMS<sup>TM</sup>. These measurement errors include local CD variations due to wafer non-flatness and processing (CD-SEM), local mask CD variations (AIMS<sup>TM</sup>) and cumulative measurement errors from the calculation method of the placement error (adding line and space measurements). To arrive at these results we fitted the original measurement curves with a 2<sup>nd</sup> order polynomial fit to get rid of most of the measurement noise, centered best focus for 1:1 comparison and calibrated the central space (SGSG) size, which is the reference for the feature-center calculation.

The feature-center placement error has also been determined across wafer and across field for the binary 40nm and 43nm modules, as shown in Figure 15. The mean placement errors (PE) for the 40nm module are equal to the BF values found in Figure 14. The variation of the placement errors across the wafer are below 2.0nm 3 $\sigma$  and across the field below 1.0nm 3 $\sigma$ . This is in line with the 120nm 8 $\sigma$  focus budget as mentioned before: from the focus sensitivity of the SG placement error an AW PE variation of 1.5nm 3 $\sigma$  is expected. From the AIMS<sup>TM</sup> measurements we can derive that 0.4nm 3 $\sigma$  placement error variation is coming from the mask. However, it was not possible to find a correlation between placement errors of individual points on the mask and the wafer because of intra field dose and focus variations on the wafer (affecting the placement error) and the measurement noise. The 43nm half pitch module shows smaller placement errors than the 40nm module, but similar across wafer and across field PE variations.

Figure 16 shows the PE for 38nm half pitch design rule. The mean placement errors are larger than for the 40nm module, making the necessity for an alternative design with reduced feature-center shift even larger. The AF PE variations are similar to the 40nm and 43nm case, which means that the focus control of the scanner is sufficient to keep these PE variations well under control.

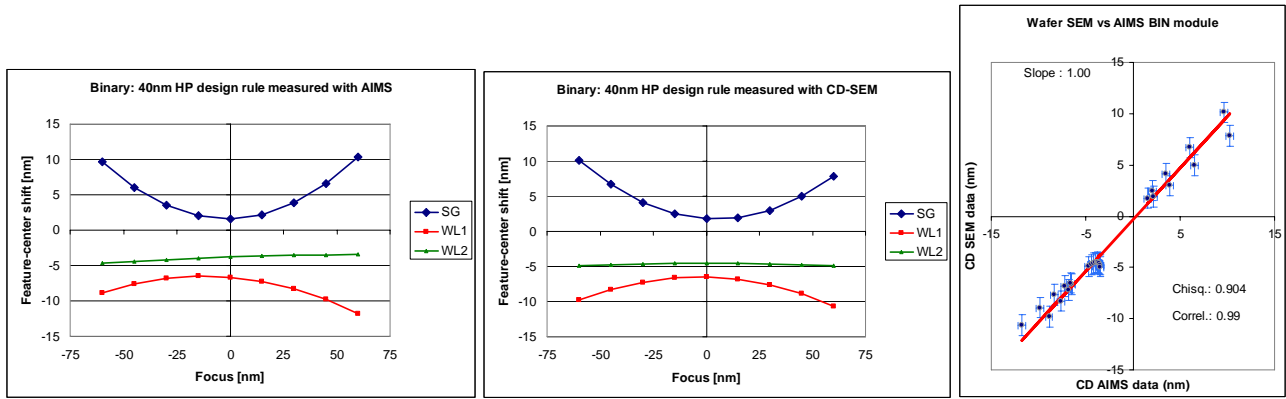


Figure 14 Feature-center placement error measured for SG, WL1 and WL2 for the binary 40nm half pitch flash module. Left: AIMS<sup>TM</sup> measurements on mask level (numbers are on 1x), Middle: CD-SEM measurements on the wafer, Right: Correlation plot of the measured feature-center placement errors between wafer (CD-SEM) and mask (AIMS<sup>TM</sup>)

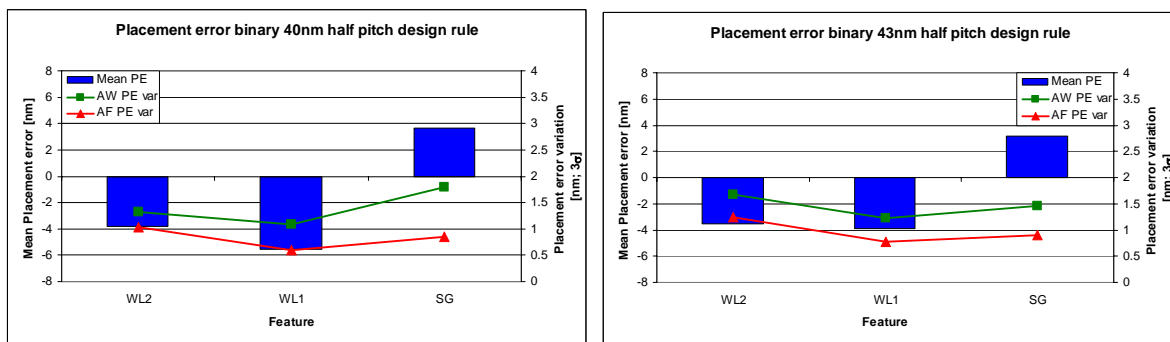


Figure 15 Mean Placement Error (PE), across wafer (AW) PE variation and across field (AF) PE variation for WL1, WL2 and the select gates as measured with CD-SEM on the wafer. Left: 40nm data, Right: 43nm data.

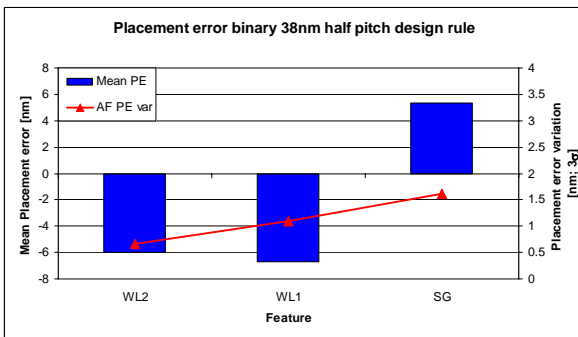


Figure 16 Mean Placement Error (PE) and across field (AF) PE variation for WL1, WL2 and the select gates as measured with CD-SEM on the wafer for 38nm half pitch design rule

### 3.6. De-sensitizing the Flash wordline mask pattern

One of the possibilities to de-sensitize the flash wordline mask pattern for feature-center placement errors is by varying the space between the select gates and the first wordline. Two designs with 1.5 and 2.0 times the design rule half pitch for SP0 have been studied in resist for the 39, 40 and 43nm modules. The 38nm module with SP0 2.0 design showed

severe pattern collapse of WL1 when illuminated with extremely high sigma settings and could not be used for analysis. The OPC treatment on the mask was done for a different illumination setting (see also Table 1), which explains this pattern failure.

Figure 17 shows the feature-center shift of SG, WL1 and WL2 for both designs of the attenuated and binary 39, 40 and 43nm modules. Again we fitted the original measurement curves with a 2<sup>nd</sup> order polynomial fit to get rid of most of the measurement noise. An increase of SP0 results in a reduction of the feature-center shift for wordline 1, but the placement error of wordline 2 and the select gates increases slightly. However, for the attenuated module the focus sensitivity of the SG shift reduces as well. So if we allocate again the full 20% of the CD node to the placement errors, this design gives the largest usable focus range: 110nm focus range for the 2.0 design vs. 50nm focus range for the 39nm half pitch 1.5 design. For the binary module this is: 100nm for the 2.0 design vs. 80nm for the 1.5 design. The

effects observed for the 39nm modules are more pronounced for the 40nm modules: The increase of space SP0 results also in a significant reduction of the feature-center shift for wordline 1, while the placement error of WL2 and SG increases. In the attenuated case the focus sensitivity of the SG shift is almost completely removed. The focus sensitivity of the WL1 shift is reduced as well, while it remains the same for WL2. This results in a usable focus range of 140nm for the 2.0 design, vs. only 70nm focus range for the 1.5 design. For the binary module the focus sensitivities are slightly reduced by increasing SP0, which result in: 120nm focus range for the 2.0 design (if we extrapolate the asymmetric curve for SG), vs. 90nm focus range for the 1.5 design. For the attenuated 43nm module such a comparison cannot be made due to missing SEM data. The allowable focus range for the SP0 1.5 design is 110nm. For the binary modules a significant improved is observed when going to the SP0 2.0 design: 130nm focus range for the 1.5 design versus 160nm for the 2.0 design.

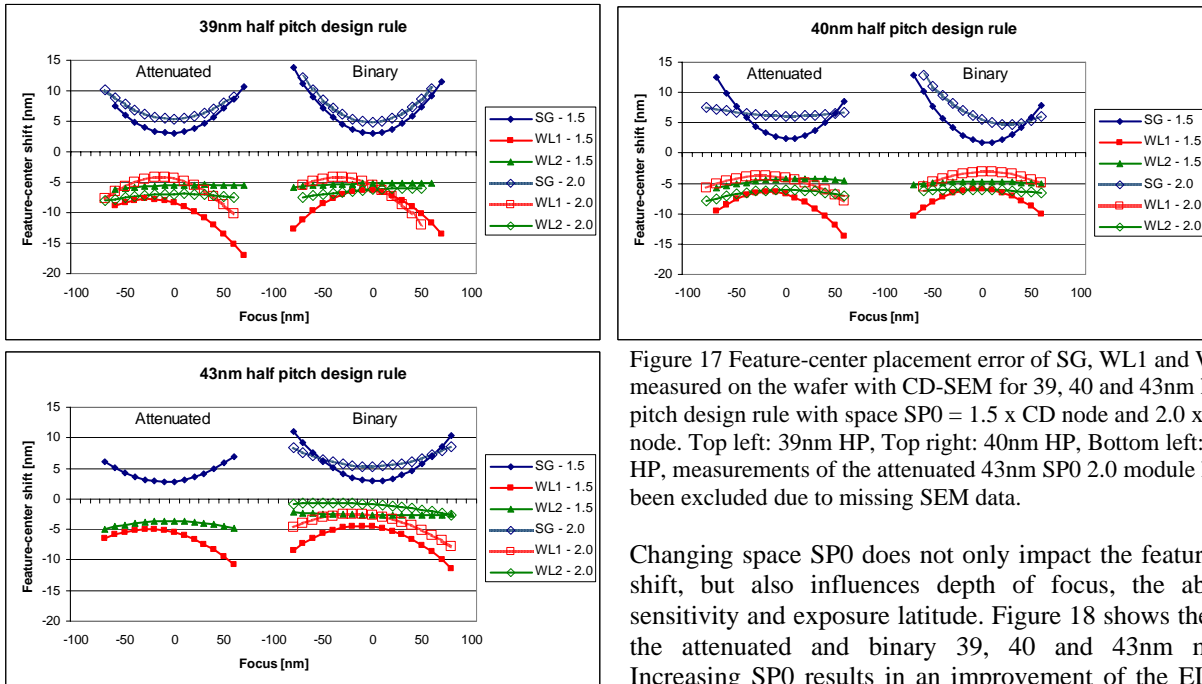


Figure 17 Feature-center placement error of SG, WL1 and WL2 measured on the wafer with CD-SEM for 39, 40 and 43nm half pitch design rule with space  $SP0 = 1.5 \times CD$  node and  $2.0 \times CD$  node. Top left: 39nm HP, Top right: 40nm HP, Bottom left: 43nm HP, measurements of the attenuated 43nm SP0 2.0 module have been excluded due to missing SEM data.

Changing space SP0 does not only impact the feature-center shift, but also influences depth of focus, the aberration sensitivity and exposure latitude. Figure 18 shows the EL for the attenuated and binary 39, 40 and 43nm modules. Increasing SP0 results in an improvement of the EL for all dose-critical features in the gate layer of the attenuated

module. This can be attributed to the change in OPC needed to print all features to size. The improvement is not observed for the 39 and 40nm binary module, which is as expected since the EL of binary features is less sensitive for line biasing, as also observed in a previous study<sup>6</sup>. However, the 43nm binary module also shows a significant improvement of the EL for the 2.0 design. The large increase of the EL of SP0 itself is misleading since the EL was calculated based on the design CD, which is larger for the 2.0 design than for the 1.5 design.

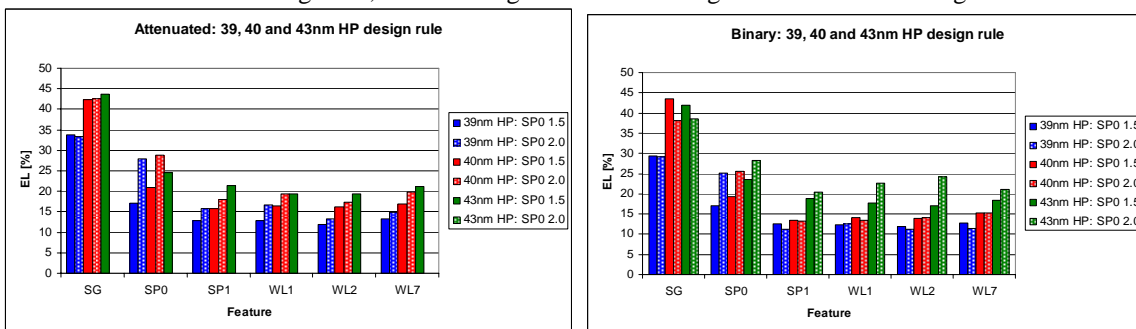


Figure 18 Exposure Latitude for all evaluated features in the gate layer for 39, 40 and 43nm design rule half pitch with space  $SP0 = 1.5 \times CD$  node and  $2.0 \times CD$  node as measured with CD-SEM on the wafer. Left: Attenuated module, measurements of the attenuated 43nm SP0 2.0 module have been excluded due to missing SEM data, Right: Binary module

SP0 is particularly sensitive for odd aberrations like coma, which result in Left-Right CD differences. Figure 19 shows the measured sensitivity for low order coma ( $Z7$ ) for the attenuated and binary 39 and 40nm modules. The  $Z7$  sensitivity has been determined for WL1, SP0 and SG by introducing a 7nm  $Z7$  tilt across the slit and measuring the L-R CD slit profile. It should be noted that also higher order coma terms, like  $Z14$  and  $Z23$  were introduced, which have been omitted from the analysis. Therefore the absolute values of the sensitivities presented here should be treated as indicative, however the trends observed still hold. Increasing SP0 results in a large reduction of the  $Z7$  sensitivity for SP0. The L-R asymmetry for the select gates also decreases. In both cases the L-R CD slit profile of the SP0 2.0 design shows hardly any tilt anymore, reducing the calculated sensitivity, but also the correlation coefficient (other contributors start to dominate the L-R CD slit profile). However, increasing SP0 does not have a positive effect on the  $Z7$  sensitivity of WL1, which even increases slightly. Figure 19 also shows that the  $Z7$  sensitivities for the binary modules are slight less (roughly 30% less) than for the attenuated modules.

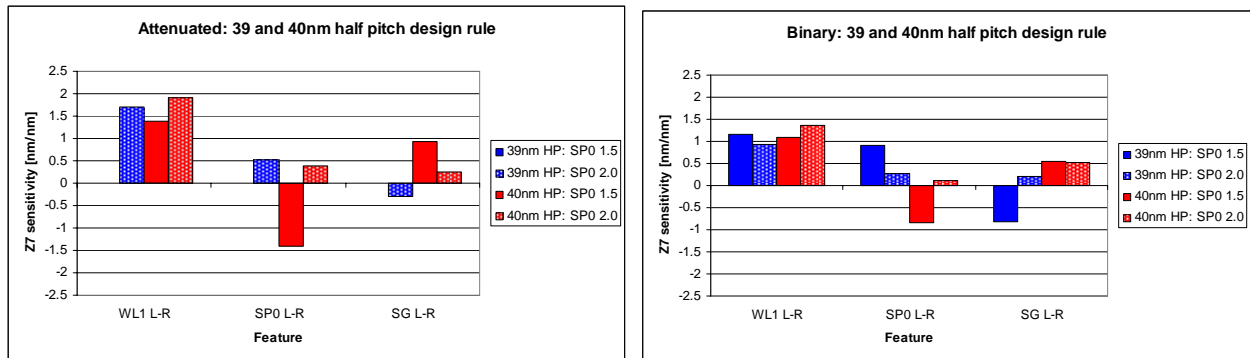
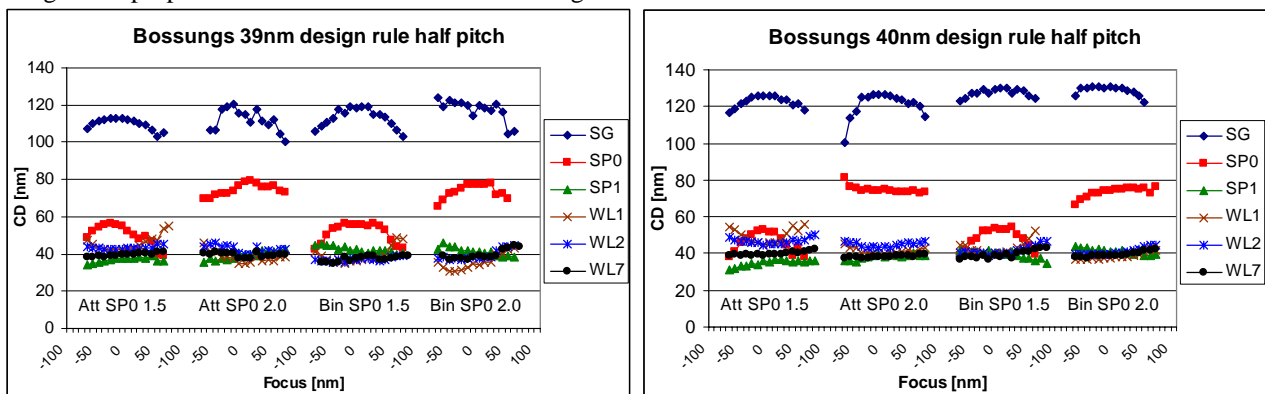


Figure 19 Measured sensitivity for low order coma ( $Z7$ ) for WL1, SP0 and SG with 39 and 40nm design rule half pitch. The SP0 1.5 design is compared to SP0 2.0 design. Left: Attenuated module, data for 39nm HP: SP0 1.5 has been excluded due to missing SEM data, Right: Binary module

Figure 20 shows the Bossung plots for the attenuated and binary 39, 40 and 43nm modules with 1.5 and 2.0 SP0 design. The increase of SP0 affects especially the Bossung curvature of SP0 itself by removing the focus sensitivity almost completely (SP0 prints close to isofocal) for both the attenuated and binary modules. But also the focus sensitivity of WL1 is significantly reduced. This is not very clearly visible in the 39nm Bossungs which are quite noisy. However it is clear that this optimization strategy is valid for multiple pitches.

Figure 21 shows CD-SEM images of the attenuated 39 and binary 38nm half pitch design rule flash cell. The 39nm module has the SP0 2.0 design, which results in a good image integrity over (almost) 130nm focus range. The 38nm module has the SP0 1.5 design, and shows a good image integrity over 90nm focus range. It is expected that the 2.0 design with proper OPC would increase the focus range for the 38nm modules as well.



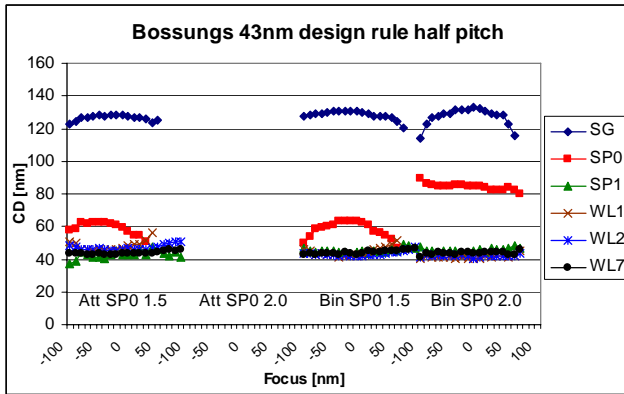


Figure 20 Bossung plots for all evaluated features in the gate layer (39, 40 and 43nm half pitch) with space SP0 = 1.5 x CD node and 2.0 x CD node as measured with CD-SEM on the wafer. Top left: 39nm HP, Top right: 40nm HP, Bottom left: 43nm HP, measurements of the attenuated 43nm SP0 2.0 module have been excluded due to missing SEM data.

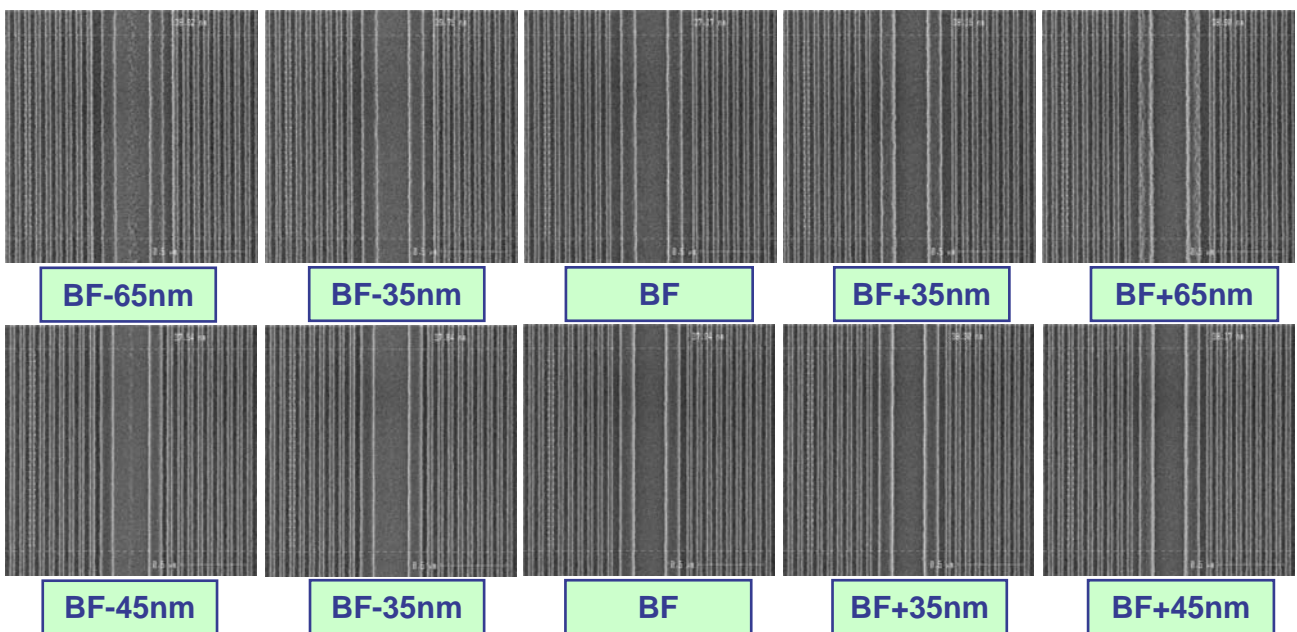


Figure 21 NAND-type flash memory gate layer exposed on an XT:1900Gi immersion scanner. Top: Attenuated 39nm half pitch design rule with (optimized) SP0 2.0 design, good image integrity is observed over 130nm focus range. Bottom: Binary 38nm half pitch design rule with non-optimal SP0 1.5 design, good image integrity is observed over 90nm focus range.

#### 4. Conclusions

An experimental study of the resolution limit of the NAND-Flash Memory Gate layer for a production-worthy process on the XT:1900Gi has shown the most important lithographic parameters for 38, 39, 40 and 43nm half pitch design rule. Full wafer CD uniformity measurements show that the requirements for the 43 and 40nm node can be met, especially if the mask design is optimized for DoF. Process window analysis and MEEF measurements indicate that there is sufficient margin for a stable process. However, feature-center placement errors consume a large part of the overlay budget if the mask design is not properly optimized. Pushing the resolution limit further down to 39 and even 38nm half pitch shows potential for a production-worthy process, although an increased MEEF and increased feature-center placement errors pose even more stringent requirements on the mask design and uniformity. Feature-center placement errors for the select gates and first few adjacent wordlines have been measured on mask level, clearly demonstrating that this intra-module overlay error is inherent to the design of the flash wordline mask pattern.

By comparing AIMS™ measurements to wafer prints using an XT:1900Gi scanner it has been experimentally proven with an independent measurement technique that feature-center placement errors in Flash memory designs are optical induced effects. It has been shown that AIMS™ measurements can accurately capture these optically induced mask effects and therefore help to distinguish between mask effects and scanner or resist effects in lithography. The feature-center placement errors can be accurately measured by AIMS™ and mapped almost perfectly on wafer measurements with CD-SEM. The AIMS™ 45-193i aerial image measurement system has proved to be a valuable tool to characterize the imaging performance of Flash memory masks in terms of CD variation, and was used to apply reticle error correction. A recent publication<sup>7</sup> shows furthermore that important lithographic parameters such as MEEF, OPE, EL, and DOF could already be quantified on mask level before the mask is actually exposed on a scanner.

We showed that the flash wordline mask pattern can be de-sensitized for placement errors by increasing the space between the select gates and wordline 1. This also results in a reduced focus sensitivity of SP0 and WL1 and lower sensitivity for odd (coma) aberrations for SP0 and the select gates. If we consider the focus-, dose-, Z7 sensitivity and the placement errors, the trends observed during this study can be summarized as follows:

	Focus sens	Dose sens	Z7 sens	Placement Error
SP0 1.5 ATT	☹	☺	☹	☹
SP0 2.0 ATT	☺	☺	☺	☺
SP0 1.5 BIN	☹	☺	☹	☹
SP0 2.0 BIN	☺	☺	☺	☺

So, overall a binary mask with SP0 2.0 design seems to give the best results in terms of CD uniformity and intra-module overlay errors, but the differences are subtle and change with the exact OPC applied.

Good image integrity over 130nm focus range has been demonstrated for 39nm half pitch design rule with SP0 2.0 design, while good image integrity was observed over 90nm focus range for 38nm half pitch design rule with SP0 1.5 design, both exposed with an XT:1900Gi hyper-NA immersion system. This shows the viability of the XT:1900Gi to support the demands of the Flash memory market for 40nm half pitch for high volume manufacturing, with potential for extension to 38nm half pitch with proper (mask design) optimization.

## 5. Acknowledgements

The authors would like to thank Carl Zeiss SMS for supporting the AIMS™ measurements, Brid Connolly from AMTC-Toppan for the support, technical suggestions and the effort to get the test mask delivered on very short notice. We furthermore would like to acknowledge the '1900 apps team' for carrying out the exposures on the XT:1900Gi, a special thanks to Mariette Berende-Hoogendijk and other members of the CD-SEM metrology group for the CD-SEM measurements and pictures.

## REFERENCES

- <sup>1</sup> J. de Klerk et al, "Performance of a 1.35NA ArF immersion lithography system for 40nm applications", Optical Microlithography XX, Proceedings of SPIE, Vol. 6520, SPIE (2007)
- <sup>2</sup> M. Dusa et al, "An integrated lithography concept with application on 45nm ½ pitch flash memory devices", Optical Microlithography XIX, Proceedings of SPIE, Vol. 6154 (2006)
- <sup>3</sup> A. Zibold et al, "First results for hyper NA scanner emulation from AIMS™ 45-193i", Proceedings of SPIE, Vol. 6283, 628312 (2006)
- <sup>4</sup> P. De Bisschop et al, "Using the AIMS 45-193i for hyper-NA imaging applications", Proceedings of SPIE, Vol. 6730, 67301G (2007)
- <sup>5</sup> E. Poortinga et al, "Investigation of hyper-NA scanner emulation for photomask CDU performance", Proceedings of SPIE, Vol. 6533, 653307 (2007)
- <sup>6</sup> E. van Setten et al, "Masks for Flash Memory Gates for the 45nm node: Binary or attenuated?", Proceedings of SPIE, Vol. 6533, 65330L (2007)
- <sup>7</sup> E. van Setten et al, "Characterizing the imaging performance of Flash Memory masks using AIMS™", to be published in Proceedings of SPIE, (2008)