SYSTEM CONSIDERATIONS FOR MASKLESS LITHOGRAPHY

T. Karnowski, D. Joy, L. Allard, L.Clonts
Oak Ridge National Laboratory, Oak Ridge TN USA 37831

ABSTRACT

Lithographic processes for printing device structures on integrated circuits (ICs) are the fundamental technology behind Moore's law. Next-generation techniques like maskless lithography or ML2 have the advantage that the long, tedious and expensive process of fabricating a unique mask for the manufactured chip is not necessary. However, there are some rather daunting problems with establishing ML2 as a viable commercial technology. The data rate necessary for ML2 to be competitive in manufacturing is not feasible with technology in the near future. There is also doubt that the competing technologies for the writing mechanisms and corresponding photoresist (or analogous medium) will be able to accurately produce the desired patterns necessary to produce multi-layer semiconductor devices. In this work, we model the maskless printing system from a signal processing point of view, utilizing image processing algorithms and concepts to study the effects of various real-world constraints and their implications for a ML2 system. The ML2 elements are discrete devices, and it is doubtful that their motion can be controlled to the level where a one-for-one element to exposed pixel relationship is allowable. Some level of sub-element resolution can be achieved with gray scale levels, but with the highly integrated manufacturing practices required to achieve massive parallelism, the most effective elements will be simple on-off switches that fire a fixed level of energy at the target medium. Consequently gray-scale level devices are likely not an option. Another problem with highly integrated manufacturing methods is device uniformity. Consequently, we analyze the redundant scanning array concept (RSA) conceived by Berglund et al. which can defeat many of these problems. We determine some basic equations governing its application and we focus on applying the technique to an array of low-energy electron emitters. Using the results of Monte Carlo simulations on electron beam profiles, we determine an empirical "impulse response" for each emitter and thus determine how each emission manifests itself in the final printed lithographic pattern. We apply methods to determine the best printable image for a variety of RSA geometries, including different levels of redundancy and achieved printer element spacing. We use concepts of total printing error to help quantify the printing quality. Through simulation, we report the effects of dead or missing elements. We also present some error analysis to account for non-ideal array positioning. Ultimately, we believe that printing quality should be the grounds for determining the necessary data rates to support competitive manufacturing with ML2 devices.

1. INTRODUCTION

Gordon Moore's law (i.e., the number of transistors on a microchip will double every 18 months) has become a driver for the semiconductor industry. Of the candidates to replace (or supplement) optical lithography, electron-beam direct-write (EBDW) technology provides the highest potential for maskless lithography (ML2). The International Technology Roadmap for Semiconductors (ITRS) states "Breakthroughs in direct-write technologies that achieve high throughput would be a significant paradigm shift [that] would eliminate the need for masks, offering inherent cost and cycle-time reductions."[i] However, there are some rather daunting problems with establishing ML2 as a viable commercial technology. The data rate necessary for ML2 to be competitive in manufacturing is estimated to be in the 1-10 terabit / s range [ii] that is not feasible with technology in the near future. There is also doubt that the competing technologies for the writing mechanisms and corresponding photoresist (or analogous medium) will be able to accurately produce the desired patterns necessary to produce multi-layer semiconductor devices [iii]. There have been several studies conducted on the lossless compression of integrated circuit patterns and the circuitry to implement them [ii,iv,v,vi]. Our work here starts from a more fundamental approach, of first investigating the printability of these patterns, with a specific eye toward a redundant scanning array (RSA) geometry using low-energy electron beams, with the ultimate goal (not addressed here) of signal compressibility and print quality.

2. SYSTEM MODELING

We explored the geometry issues associated with a tilted redundant scanning array or RSA proposed by Berglund [vii]. The RSA concept is illustrated in Figure 1. In Figure 1a, we see an array with uniform spacing in the x- and y-directions. As the array moves downward in the y direction, we see that each site at position $y = y_a$ is hit by five array elements. In this case, we say the array has created a *redundancy* of 5. In Figure 1b, we tilt the array by 45 degrees. As a result, since we have not increased the number of array elements, we decrease the available redundancy to range from 5 to 1. However, we have decreased the effective spacing of the x-positions at $y = y_a$. Figure 1c shows in detail how we use the array tilt to create a new effective sub-array of elements at smaller inner-element spacing than a non-tilted geometry.

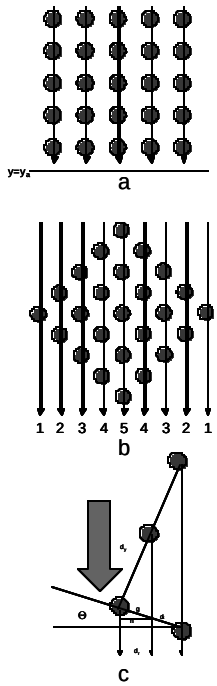


Figure 1 Illustration of redundant scanning array (RSA). a) shows the effect of redundancy, with 5 elements crossing the y = ya line at the same place. b) shows that tilting the array can create smaller inner-element distances, and c) details this fact

Our analysis revealed that there are only discrete angles that can achieve co-incidence. These angles obey the equation

$$\tan(\Theta) = -k\frac{M}{N} \qquad \textbf{Equation 1}$$

where M and N are integers and k is the ratio of the element spacing in the x and y direction. Essentially, a ratio of M=1 and N = 5 means co-incident elements are 1 column apart and 5 rows back. For M=1, N is approximately equal to

$$N = \frac{d_{element}}{d_r} \qquad \textbf{Equation 2}$$

where $d_{element}$ is the spacing between array elements (assumed to be identical in the x and y direction, or k=1) and d_r is the distance between steps or array "firings". Assuming M=1, as N >> M we find that the distance between co-incident elements becomes approximately

$$d_{coincident} = Nd_{element} \qquad \textbf{Equation 3}$$

or

$$d_{coincident} = \frac{d_{element}^2}{d_r} \qquad \textbf{Equation 4}$$

Finally, as the array moves the number of steps required between co-incident elements is

$$steps = \frac{d_{element}^2}{d_r^2} \qquad \textbf{Equation 5}$$

Due to the 2-dimensional nature of the problem the number of steps scales as the square of the ratio between the actual spacing and the desired spacing. Thus, although the tilted array is an ingenious idea with some advantages (particularly that motion is only required in one direction), it cannot get around the fact that many steps are required to write a tight pattern with relatively sparsely spaced emitters.

The redundant scanning is also an ingenious idea and very necessary to overcome the shortcomings of binary emitters, but the number of steps required scales linearly with the number of redundant elements desired.

As a result, a step-and-settle system will likely be impossible unless the emitters can be constructed extremely close together. With a redundancy of 15, a ratio of $d_{element}/d_r$ of 100 results in 150,000 total steps, or (with a target of 1 wafer per minute) 2500 steps per second. Achieving rates near 100 steps per second may be possible circa 2006, but these rates are not likely to increase by orders of magnitude in the foreseeable future. Emission technologies that depend on step-and-settle will probably not be feasible.

2.1. Linear System With Threshold

To determine the location of the firing sites, we first construct an array given a particular d_x and d_y. The choice of these values will depend on the manufacturing methods for the maskless writer array. The firing site density is related to the impulse response; since the beam width is assumed to be less than 25 nm total, we use effective array spacing that are no greater than approximately ½ this distance (about 10 nm). We then estimated the size of the array required for a given redundancy. We actually employed several different values and analyzed their effect on data rates and error rates; however, we did not consider the bad element compensation capability of the redundancy. We then created an array with the spatial coordinates of the rectangular emitter array. Our next step was to rotate the array element locations from old coordinate system (x',y') to new coordinates (x,y). When the tilting was performed, several elements contributed to x locations with insufficient redundancy; and several contributed with excessive redundancy. We filtered out these elements, and then proceeded through a timed sequence of stepping the array by d_r in the y direction only. We proceeded for a number of discrete time steps. We thus generated a table of values that show the position of each writer as a function of time. This table had both a true, real position and a truncated integer position as well. We required the integer positioning because our analysis software was based on image processing, and thus pixels could be off-on only and were discrete positions.

The overall system is modeled as shown in Figure 2. Discrete "firing" points are found based on the array geometry and step or firing size. Each weight ranges from 0 (no firings) to R, where R is the chosen redundancy. Once these are found, the system impulse response is applied through a convolutional process. This assumes that the system is linear; exposing a site with three signals of strength "X" is equivalent to exposing the site with a single signal of strength 3X. Finally, we model the development process with a simple threshold. This does not account for things like line edge roughness or variability in the development process, but it does help us model the effects of the impulse response coupled with the selected printer weights.

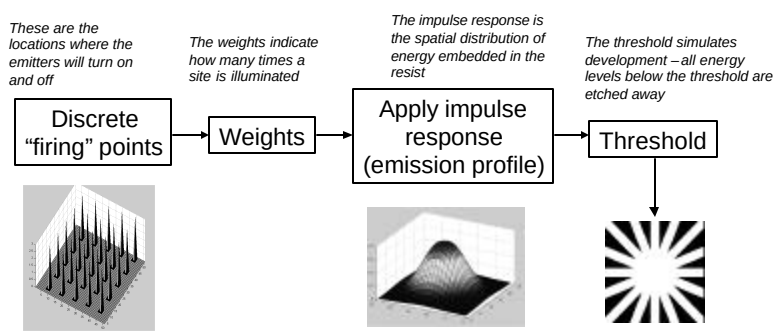


Figure 2 Maskless lithography system concept

2.2. Resist Response

We seek to use multiple gray levels to allow fine positioning of energy deposition. Multiple exposures of a site on the resist media creates different degrees of modification that can be removed by the development process. As a result more precise positioning of the printed pattern can be achieved.

It is difficult to determine a general rule for resist responses to low energy e-beam deposition. Therefore, this section will present the results of Monte-Carlo simulations performed under the direction of David Joy [viii]. We will then show how we used these results to determine an impulse response suitable for linear systems analysis for our printing simulations.

The media response simulations used photoresist parameters similar to PMMA under low-energy e-beam conditions. In the simulations a specific beam profile was assumed, with a 25 nm profile. Figure 3 shows the results of the simulation

with delivery into 10 nm of a PMMA type material, as well as an idealized response. The profile is normalized to a 100% full saturation dose or FSD. A 100% FSD means the material is exposed to saturation such that further exposure results in no change in the chemical composition of the resist media. Thus, 99% is practically the full dose, and 10% is 1/10th of FSD. Note that the actual "spot" produced by these beam exposures is twice the size shown, since the spot is symmetric about the beam axis.

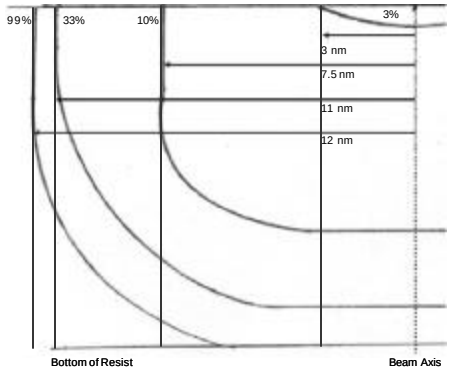


Figure 3 Resist response to 25 nm, 1 kEV beam into 10 nm PMMA material. Black lines represent simulation results; Red lines represent idealized results used for impulse response.

As noted, this assumes delivery into a 10 nm of PMMA, which presents two problems. First, such a thin deposition of resist media is difficult to achieve in practice. Second, with this configuration the depth of the developed and exposed region does not reach the full 10 nm until almost 99% full exposure. Such a system is effectively useless until full saturation is achieved, which means it is difficult to attain gray scale levels. We attempted to investigate alternate resist media, including chemically amplified resists, but we were not able to complete simulations of these materials due to time and budget constraints. However, for our modeling, we assumed that we can obtain a resist material that does not have this problem. We therefore assumed that we could obtain a resist material with a response more like that of the red lines in Figure 3, where we acquire a dosage width similar to our PMMA simulations but with a depth (possibly attained through chemical amplification) that reaches the bottom of the resist thickness.

We took these results and extrapolated them using a second order exponential profile. We modified the profile so that it did not become unbounded at the origin. We eventually found the equation

$$h(r) = \begin{cases} 1.8\frac{G}{R}, r \le 1 \\ \frac{G}{R}\exp\left(-0.0283r^2 + 0.0613r - 1.0546\right) elsewhere \end{cases}$$ **Equation 6**

where G is the single dose gain, R is the redundancy, and r is the distance from the center of the emitter or

$$r = \sqrt{x^2 + y^2}$$ **Equation 7**

The profile is illustrated in Figure 4. In Figure 5, we show the resulting spot width after threshold development with the threshold set to unity for G=1, R=1, which means a single dose is equivalent to 1% of a full saturation dose. In Figure 6, we show the resulting spot width after threshold development for G=5 and R=3 for comparison. Note that in the case of 1 dose for R=3, the energy deposition is so low that no development occurs. However, these low doses allow us to gain a greater level of control and shaping over the other levels that can be achieved.

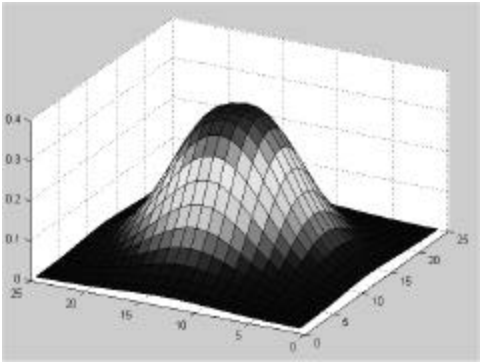


Figure 4 Two-dimensional profile of impulse response

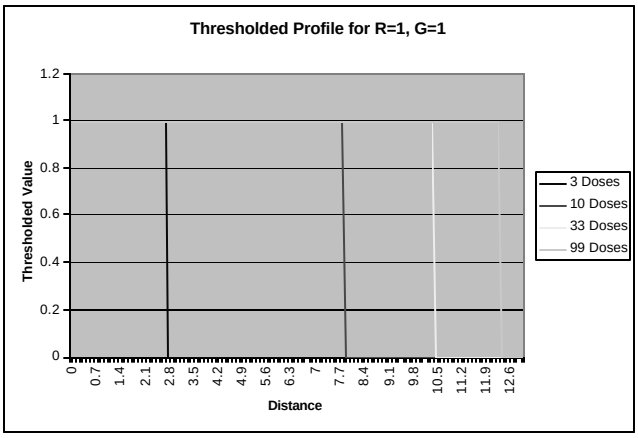


Figure 5 Thresholded profile for R=1, G=1 (where 1 dose = 1% of FSD)

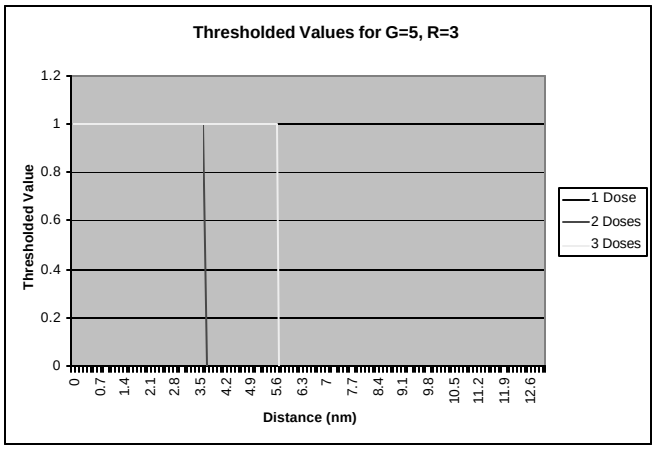


Figure 6 Thresholded profile for G=5, R=3

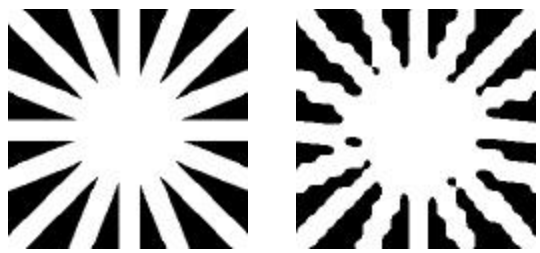
3. DETERMINING FIRING WEIGHTS

Our next step is to find the optimal printing patterns for a particular pattern. Essentially, we can locate particular "firing sites" on a printing field, which represent positions where the emitter elements will energize the resist medium. Each firing site will be assigned a gray level ranging from 0 to R-1, where R is the redundancy. These levels are the number of times an emitter passing over the site will switch on or off. Pseudoinverse filtering and Weiner filtering are operations which are designed for solving this type of problem. However, this situation is complicated by two factors: the quantization of the weights to integer values from 0 to R, and the thresholding operation which is non-linear and difficult to model.

Our error was computed by simply taking the 1-norm (absolute difference) between the desired pattern and the resulting thresholded pattern, with a penalty of 1 for any pixel where the test result does not match the desired pattern. This calculation does not attempt to increase the penalty on more obvious problems, such as bridging between structures, but nor does it soften less painful issues such as slight pixel errors on long lines.

$$Error = \sum_{y}\sum_{x} abs\left(P_d(x,y) - P_p(x,y)\right) \qquad \textbf{Equation 8}$$

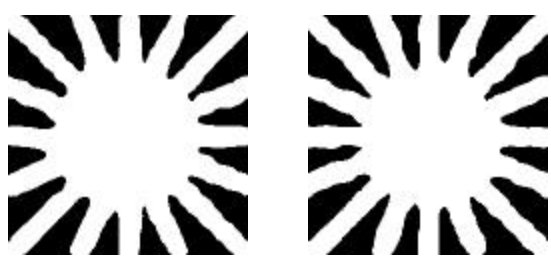
If we simply sample the desired pattern, the result is corrupted by aliasing. Figure 7 shows an example of this operation, with the original pattern and the output signal. This approach is reasonable for a first approximation but delivers very poor results. We used four gray levels (R=3) and an element spacing of 10 pixels for these experiments.



a b

Figure 7 Results of omitting the analysis filter G. (a) input pattern; (b) output print. The error in this case is 4320 pixels, or approximately 8% of the total

Better results are obtained by filtering the pattern prior to sampling. We show results of both a band-limiting filter and also using the impulse response itself in Figure 8. In both cases we take the desired pattern, apply the filtering process, sample it with the firing array, round the values to the closest integer ranging from 0 to R, then use these values to simulate the print by convolving with the emitter impulse response and thresholding. In this mode, prior to thresholding we ramped through several different scalings and selected the one that gave the least error. Both these techniques yield better results than the unfiltered case.



a b

Figure 8 Direct method of weight synthesis (a) filtered with impulse response; (b) filtered with anti-aliasing filter. The error is 2831, or about 5.31%, and 2543 pixels, or about 4.76%

A special case of filtering prior to sampling is to subsample the desired pattern and sum up all the samples in the vicinity of a sampling site, then assign the weight based on this value. This technique yields good results quickly but does not necessarily yield the best results nor the most robust, as seen in Figure 9.

Figure 9 Direct method of weight synthesis using sum of values around firing sites. The error is 3573 pixels, or 12% of the total.

By writing the process of the convolution between the impulse response and the firing sites as a system of equations, we can create a problem that can be solved using a least-squares approach. The equation of interest is

$$V = HF \qquad \textbf{Equation 9}$$

where V,H, and F are matrices:

V is the output (pixels x 1): it is the output energy deposition function. After thresholding it should closely match our desired image. V can assume any non-negative value. We set the value of V to 0 for areas where we do not want pattern and we set it to T for areas where we do want pattern.
H is the transformation (pixels x firing sites): this transfer function tells us how a firing site location affects the output image pixels. It is a function of our image size, firing site geometry, and response of the resist to the emitter signal. H is nonnegative for the ML2 printing problem.
F is the firing weights (firing sites x 1): this matrix shows the desired firing site weight for each firing site location. We seek to solve for this value, subject to the constraint that the values of F must be discrete values from 0 to R, where R is the redundancy.

Solving this equation with a least-squares approach consisted of computing the pseudo inverse

$$F = \left(H^T H\right)^{-1} H^T V \qquad \textbf{Equation 10}$$

then truncating F to the nearest discrete value in the range {0,R}, and thresholding the result. We get a solution such as that shown in Figure 10. While our result F is guaranteed to be the least-squares values of firing sites that will most closely match V, there are two issues with this approach. First, we do not know what the value of V should actually be; there are an infinite number of possible Vs that could be thresholded and produce the desired output pattern. Secondly, the truncation and rounding of F is a major problem that is not reflected in this solution method. These issues bear further investigation to make this method useful.

Figure 10 Results using LSQ algorithm. The error in this case was 3062, which is about 5.5% error

Finally, we utilized a technique we have coined "Direct Quantized Search" (DQS) derived from the "Direct Binary Search" method [ix]for computer generated holography and image half-toning. In the DQS method an initial guess is performed at the ideal weights, then a particular site is cycled through all available weights and the error measured at each step. The weight with the minimum error is retained and the algorithm proceeds to another site. When the error cannot be reduced further, the algorithm either stops, or investigates sites in pair or four-wise fashion by selecting sites that were "free", in that all values impacted the end result equally. This method is rather time consuming but it does produce better expected error than the other techniques. We have refined the algorithm as well with some additional image processing steps to try and eliminate sites that have little impact on the overall image quality (i.e., sites that are obviously in the middle of an all-white or all-black area). An example of the results of this algorithm is shown in Figure 11, for a low redundancy of 3.

Figure 11 Results using DQS algorithm. The error in this case was 1634, which is about 3% error.

4. EXPERIMENTAL RESULTS

We ran a variety of small test patterns to test the performance of the DQS algorithm in finding reduced-error printing patterns. Next, we performed a Monte-Carlo style analysis on the DQS optimal printing patterns to determine their susceptibility to errors in array positioning. Finally, we studied through simulations the effect of dead or missing elements.

Our experimental array configuration used an array with element spacings of 100 "units". We computed array tilts for an effective spacing of 10 and 5, and used sample patterns that covered a 256 by 256 area. The sample patterns have a minimum "feature size" of 20 pixels. A convenient way to normalize these values is by taking the minimum feature size and expressing it as λ_6, indicating that this is the design rule we wish to print.

We conducted experiments with arrays using redundancies of 3, 7, 15, and 31. In all cases, we used a factor of 5 for the constant single dose gain. Intuitively, we were certain that higher redundancy and smaller effective spacings would produce the least printing error, but would also require the highest data rates. This turned out to be the case but the trade-offs our analysis illustrates are very valuable.

We ran a variety of test patterns shown in Figure 12. We show the results of tests with these patterns in Figure 13 through Figure 18.

a b c

Figure 12 Patterns used for testing various ML2 array geometries. a) Cross; b) Horizontal; c) Vertical. In all cases the bars that comprise the pattern are approximately 20 pixels wide.

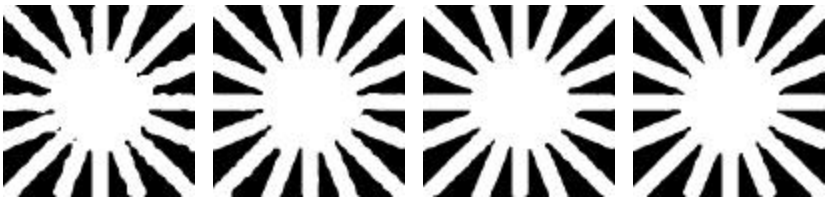


Figure 13 Results with cross pattern for spacing of 10. R=3,7,15,31

Figure 14 Results with cross pattern and spacing of 5. R=3,7,15,31

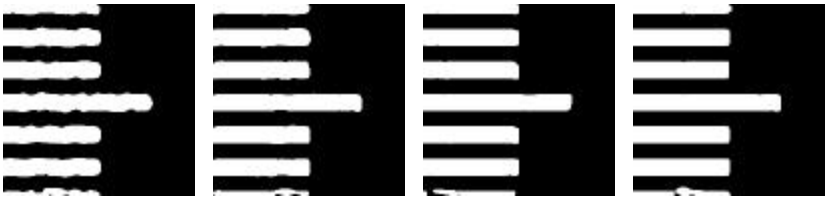


Figure 15 Results with horizontal pattern and spacing of 10. R=3,7,15,31

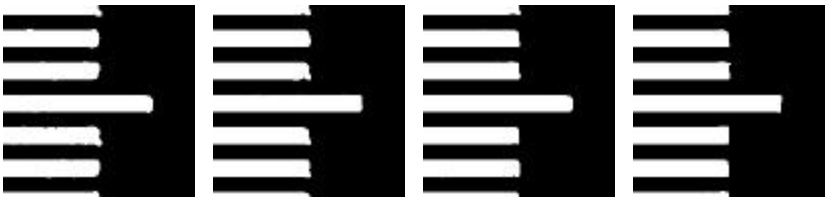


Figure 16 Results with horizontal pattern and spacing of 5. R=3,7,15,31

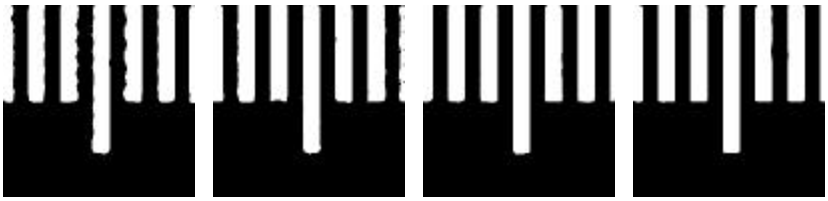


Figure 17 Results with vertical pattern and spacing of 10. R=3,7,15,31

Figure 18 Results with vertical pattern and spacing of 5. R=3,7,15,31

The positioning of the array is critical to successful writing of the patterns. Errors in the positioning will naturally occur, however, so we decided to do a Monte Carlo analysis to determine the effect of these errors. Our simulation involved skewing the positions of the writers randomly, computing the resulting pattern, and determining the error. For our analysis, we performed 16 trials for each different element spacing and redundancy. Each trial consisted of randomly perturbing the element position by generating a pair of Gaussian-distributed random values, one each for the x shift and y shift. Both had a variance of 1, which corresponds to one pixel of error. Fractional pixel shifts were computed by interpolating the impulse response function. Ideally, at time t, the random error in positioning of element N should be somewhat correlated with the error on all the other elements – but for our analysis we assumed they were all independent.

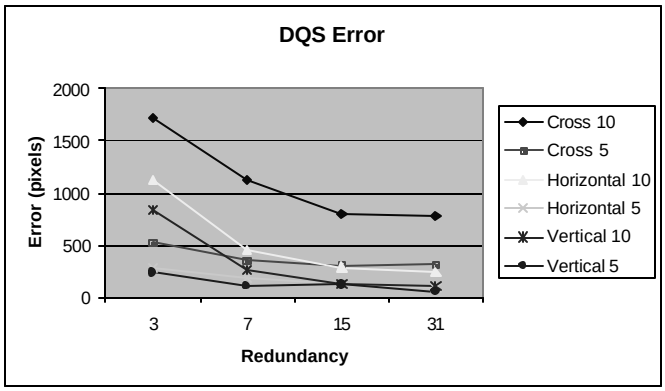


Figure 19 Plot of DQS Error vs Redundancy

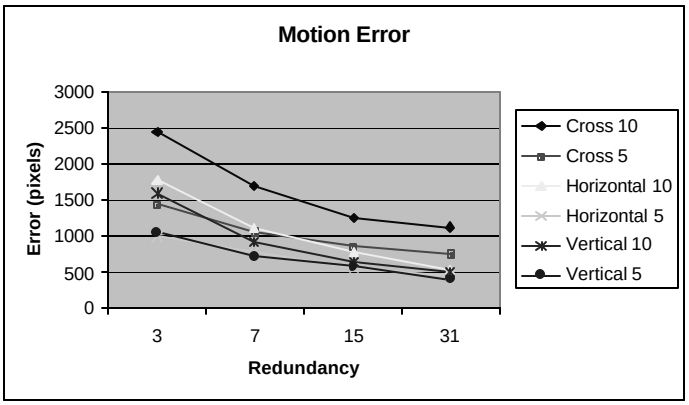


Figure 20 Plot of Motion Error vs Redundancy

We ran some experiments on the test images for two cases of redundancy and spacing that had similar data rates: 10 spacing with 15 redundancy and 5 spacing with 3 redundancy. The experiments were run by dropping a different number of random emitters for the entire writing time. Essentially, even though the writing signals were not modified

to handle known dead elements, the redundancy of 15 with larger spacing provided useful signal overlap and permitted the patterns to degrade more slowly than cases with fewer redundancy and smaller spacing.

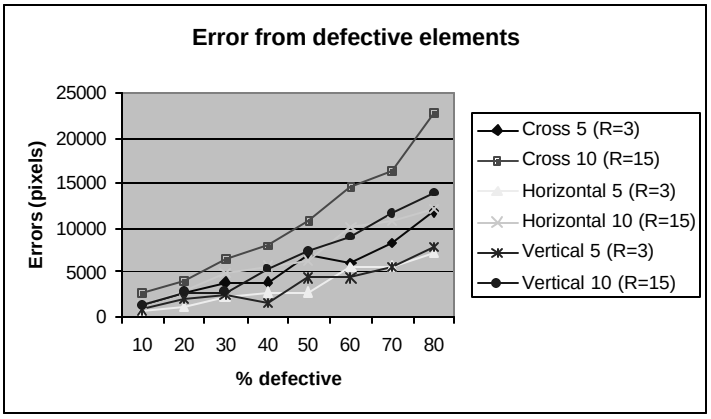


Figure 21 Plot of errors from defective elements

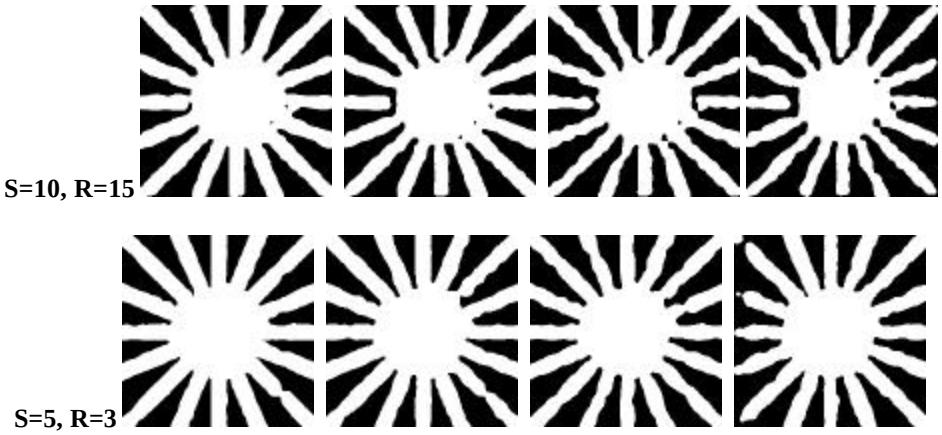


Figure 22 Cross test images showing effects of random dead elements for 10%, 20%, 30%, and 40% bad elements.

5. CONCLUSIONS

If the cost of EUV masks and other NGL technologies is excessive for small production runs, some maskless technology will prove attractive even if its throughput is slower. Increasing the allowable processing time to 30 minutes results in a data rate reduction by a factor of 30, which would achieve the "compression" required with no real work at all. Ultimately cost will be the deciding factor.

Nevertheless, there are many factors that encourage the development of very high throughput maskless lithography technologies. It is well known that large data rates are required to make maskless lithography competitive with more conventional technologies, there are other issues as well. The use of a RSA allows smaller spacing than permitted by the actual emitter spacing, but the tilted array cannot get around the fact that many writing cycles are required to write a tight pattern with relatively sparsely spaced emitters. Due to the 2-dimensional nature of the problem the number of steps scales as the square of the ratio between the actual spacing and the desired spacing. In addition, redundant scanning overcomes the shortcomings of binary emitters, but the number of steps required scales linearly with the

number of redundant elements desired. There may be other ways to construct the arrays but we have had difficulty finding a better concept; trying to directly place elements behind one another, for example, doesn't really solve the problem although it can cause coincident elements (with similar data signals) to occur more closely in time and space, possibly improving signal compression rates. Finally, due to the large number of steps, a step-and-settle system will likely be impossible unless the emitters can be constructed extremely close together. A ratio of desired emitter spacing to actual emitter spacing of 100 results in 2500 steps per second to achieve the 1 wafer per minute rate. Based on commercial development in wafer defect detection instruments that we are familiar with, achieving rates near 100 steps per second is highly difficult and may be possible around 2006, but these rates are not likely to increase by orders of magnitude in the foreseeable future. Finally, creating the optimal weighting factors will be time-consuming. This may be a fertile field for some clever solutions. We note that the problem should be somewhat easy to scale, making it well suited for networked computer solutions.

On the positive side, we believe we have shown that based on error measurements, we believe (if our simple thresholding is a reasonable practice) that a lower redundancy / gray level number can achieve effective results. Also, missing elements can be a problem, as can motion problems, but these both are not show-stoppers if the element spacing overlaps the impulse response of the system enough. We believe explorations in maskless lithography data handling should directly address the desired print quality and possibly look for solutions that can both achieve effective data rates and print quality within some tolerances.

REFERENCES

i. International Roadmap for Semiconductors, Semiconductor Industry Association, 1997.

ii. Dai, V., Zakhor, A., "Binary Lossless Layout Compression Algorithms and Architectures for Direct-write Lithography Systems", Department of Electrical Engineering and Computer Science, Univ. of California / Berkeley, 2000.

iii. Lammers, D., "EUV gains as venture ends e-beam litho work", EE Times, January 8, 2001

iv. Dai, V.; Zakhor, A., "Lossless compression techniques for maskless lithography data", Proceedings of the SPIE: Emerging Lithographic Techniques VI, V. 4688, p. 583, 2002.

v. Wild, B., *Data Handling Circuitry for Maskless Lithography Systems*, Master Thesis, UC Berkeley, 2001.

vi. Ellis, R., Kahng, A., and Zheng, Y., "JBIG Compression Algorithms for "Dummy Fill" VLSI Layout Data", UCSD Department of Computer Science and Engineering, March 13, 2003.

vii. Baylor, L.R.; Lowndes, D.H.; Simpson, M.L.; Thomas, C.E.; Guillorn, M.A.; Merkulov, V.I.; Whealton, J.H.; Ellis, E.D.; Hensley, D.K.; Melechko, A.V., "Digital electrostatic electron-beam array lithography", J. Vacuum Science & Technology B (Microelectronics and Nanometer Structures), V. 20, p. 2646, 2002.

viii. Joy, D., "The Impulse Response of a Nanolithography System", ORNL/UT-Knoxville internal report

ix. Allebach, J.P., "DBS: retrospective and future directions [direct binary search]", Proceedings of the SPIE, Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts VI, V. 4300, p 358, 2001