PLANARIZATION BY CHEMICAL MECHANICAL POLISHING: A RATE AND UNIFORMITY STUDY

A.E. Freeny
W.Y-C. Lai

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

This memorandum will be included in the volume *"Statistics in the Semiconductor Industry: Case Studies of Process/Equipment Characterization"* to be published by SEMATECH in 1993.

Planarization by chemical mechanical polishing (CMP) is gaining importance as Very Large Scale Integration (VLSI) pushes below the 0.5μm regime. The global planarity which can be achieved by this technology permits maintaining fine features with increasing metal levels. In this article we describe a designed experiment that we used to help identify an optimum processing regime and quantify process reproducibility. Looking at the effects of platen and wafer rotation on rate and uniformity of oxide removal, we discovered a tradeoff between rate (throughput) and uniformity. A process drift over time was quantified. These results contribute to a judgement of cost and timing in introducing a new process and equipment into manufacturing.

1. EXECUTIVE SUMMARY

1.1 Problem

Advanced photolithography in Very Large Scale Integration (VLSI) increasingly demands global planarity across a chip-sized printing field for fine resolution. Planarization by chemical mechanical polishing (CMP) is a simple technique that can achieve this. In a designed experiment we evaluate a polisher which does oxide planarization by CMP for possible use in the wafer fabrication manufacturing process. Our goal is to find the maximum rate of oxide removal which can be used (for throughput reasons) without degrading the uniformity of the removal over the surface of the wafer.

1.2 Solution Strategy

In CMP a wafer is held by a rotating carrier and is polished by pressing the wafer face down onto a polishing pad on a rotating platen. The important parameters for the polishing process are platen and wafer rotation frequencies and polishing pressure. Design of experiments (DOE) is an effective way to investigate the relationship of these parameters to oxide removal rate and uniformity across the wafer, while quantifying the reproducibility of the removal process from wafer to wafer. We designed an experiment to study the effects of platen and wafer rotations for an automatic wafer polishing machine suitable for CMP. All combinations of five wafer and three platen rotation frequencies were used. Three wafers were polished at each combination of wafer and platen rotation frequencies. A tendency of the polisher removal rate to drift lower through time had previously been noticed. Thus an important design issue was to choose the order of the fifteen combinations so as to allow estimation of the effects of the design parameters independent of a linear drift as well as to quantify the magnitude of the drift. Graphical methods and analyses of variance allow detection of the effects of the

parameters and estimation of the relationship of uniformity and removal rate to platen and wafer rotation frequencies.

1.3 Conclusions

Both the polishing rate and across-the-wafer non-uniformity showed large increases with increasing platen rotation and slight increases with increasing wafer rotation. A simple linear model is appropriate for quantifying these effects. A preferred operating regime for platen and wafer rotation frequencies was derived. The presence of the suspected drift was confirmed as a decrease in polishing rate over time regardless of the rotation frequencies. This drift, combined with large wafer-to-wafer variability, results in imperfect process reproducibility even with automation. This lack of reproducibility is an important factor in judging the cost and timing of introducing CMP into manufacturing.

2. PROCESS

Chemical-mechanical polishing (CMP)[1] is a novel approach to planarizing oxide topography. The procedure presses a rotating wafer face-down onto the surface of a rotating padded platen impregnated with a slurry of extremely fine abrasive. The key parameters affecting the removal rate and uniformity across the wafer are the wafer pressure against the polishing pad, the platen and wafer rotation frequencies.

The polisher used in the experiment was an automatic cassette-to-cassette loading single-wafer polisher with five stations: input, primary polish, secondary polish, rinse, and output. Wafers are removed singly from the input cassette and transported via the carrier to the primary polishing station. The polishing action consists of three relative motions: the wafer rotation about its center, the platen rotation, and a small oscillation of the arm supporting the wafer carrier. After primary polishing is completed, the carrier transports the wafer to the secondary polishing station, which performs a clean-up step. The polished wafer is unloaded via a water track into the output cassette, and kept submerged until cassette removal.

3. DATA COLLECTION PLAN

3.1 Thickness Measurements and Derived Responses

Figure 1 shows the 37 sites from the 49 possible sites where oxide film thickness was measured on experimental wafers. Each wafer was measured before and after polishing so that a comparison could be made on a site-by-site basis. In this study, measurements from sites $27-33$ and $36-40$ were eliminated, leaving 37 sites: the former because of proximity to the wafer scribe mark; the latter because of proximity to the wafer flat. Both

the scribe mark and the flat induce nonuniform oxide removal because of local topography disruption. The amount of oxide removed was calculated as the difference between the thicknesses before and after polishing at each of the 37 sites. The responses of interest were the oxide removal rate and the uniformity of the removed oxide film. The rate, R, was determined by dividing the mean of the thicknesses of the removed oxide layer measured at the 37 sites by the polishing time (three minutes in this study). The uniformity, U, was computed as the standard deviation of the thicknesses of the removed oxide layer at the 37 sites[1].

Figure 1. Map of thickness measurement locations.

1. There are two measures of the quality of the polishing process: the final planarity and the removal rate uniformity. For actual topography reduction, the final planarity is the important measure. For process evaluation, because of the number of wafers and measurements involved, planar oxide monitors with no underlying topography are used for the study. The initial uniformity of these wafers is very good, which allows us to infer the goodness of the final uniformity from the removal uniformity.

3.2 The Design of the Experiment

Previous work[1] indicated that the polisher may not reach equilibrium immediately after a change in the parameter settings. Therefore we used three wafers in each experimental condition, in case the first wafer polished behaved differently from the other two. To minimize initial wafer-to-wafer variability, all wafers came from a lot of 50 oxide-coated unpatterned wafers, which was generated in two batches of 25 wafers each. Three wafers were randomly selected from the two batches so that each polishing condition would be performed with one wafer from one batch and two wafers from the other.

The two parameters of interest were platen (polishing pad) rotational frequency, P, and wafer (wafer carrier) rotational frequency, W. Previous work[1] and other recommendation had suggested a good operating range of 10 to 20 rpm for platen frequency and 10 to 55 rpm for wafer frequency. Therefore, we selected the three values 11, 15, and 19 rpm for platen frequencies, and five values varying from 12 to 52 in increments of 10 for wafer frequencies. The 15 polishing conditions combining every platen frequency with every wafer frequency formed a full factorial experiment. Pad pressure, Pr, was held constant at a single optimum value based on previous work.

Because both P and W are quantitative, they are parameterized using orthogonal polynomials, which allows us to evaluate the shape of the functional relationship of the response to each of the parameters. Thus the effect of each parameter is seen as the sum of a linear, quadratic, etc, component; and, because of the orthogonality, we can quantify the importance of each component. The P by W interaction, can be broken into polynomial component combinations similarly[2].

Since properties of the polishing pad are likely to change with use, we were concerned that evaluation of the effects of the parameters would be affected by a drift (time dependence) of the polisher performance. There is no sequence of runs for a three by five experiment that makes estimates of effects of the parameters completely free of a linear effect of polisher drift. In this experiment, the sequence of runs was chosen to confound the effect of a linear drift with a high order component of the interaction, specifically the P quadratic by W cubic component[3]. This effect is denoted in the figures

---

2. A discussion of parameterizing quantitative factors in this way can be found in, e. g., Hicks[2].
3. "Confounding" means that we cannot tell the difference between a linear drift in the polisher and this component of the interaction. However, real high-order interactions of this type are unexpected, so we are willing to assume that the size of this component of the interaction will reflect the linear time dependence of the polisher and not some underlying physical relationship of the response to the factors.

by PQWC. The appendix describes how the run sequence was derived. The main effects of P and W, and the low-order components of the P by W interaction are essentially independent of a linear drift in the polisher. The sequence of polishing conditions is given in Table 1.

Table 1. Platen and Wafer Frequency Experiment Sequence

Run #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
P	15	19	11	15	11	19	11	15	19	11	19	15	19	11	15
W	22	42	42	52	12	12	32	32	22	52	52	12	32	22	42

It required several hours to polish the 45 wafers. The time at which each of the wafers was polished was not recorded, however the experiment ran smoothly enough so that the assumption of equally-spaced measurements in time needed to estimate the magnitude of the drift is not unreasonable. One polisher problem requiring operator intervention was noted. It occurred when polishing the third wafer at P = 11, W = 32. We discuss an anomalous result for uniformity which may be related to this interruption in Section 4.4.

4. ANALYSIS AND INTERPRETATION OF RESULTS

4.1 Graphical Presentation

The boxplots[4] in Figure 2 show the distribution of removal rates over the 37 sites on each wafer. It is organized so that W appears as rows and P as columns. The three boxes within each P, W combination are shown in the order in which the three wafers in that condition were polished. Information relating to removal rate and uniformity can be found from the boxplots by noting that both the mean (removal rate) and median are measures of location, and both the standard deviation (uniformity) and vertical size of the box are measures of spread.

Figure 2 indicates that removal rate increases as P increases. Furthermore there appears to be a lot of variability among the three wafers in any experimental condition, both in location and spread. It also seems that the boxes are larger, indicating more variability, in conditions with P = 19 rpm. The third wafer polished with P = 19 rpm and

4. A box is drawn for every wafer summarizing the distribution of the observations over the 37 sites per wafer. The large dot within the box is plotted at the median, and the box boundaries signify the upper and lower quartiles (+−25%). "Whiskers" are drawn out to the closest data point not more than 1.5 times the interquartile range from the median. Points beyond the whiskers are plotted individually.

Figure 2. Boxplots of Removal Rate in Angstroms/minute. The figure is organized so that wafer frequency (W) appears as rows and platen frequency (P) as columns, with three boxes for each P, W combination. Each box summarizes the distribution of the 37 removal rates for one wafer.

W = 22 rpm appears to be unusual in both location and spread.

4.2 Model for Analyses of Variance

To evaluate the effects of the parameters P, W, and O, the order in which wafers were polished in each experimental condition, we fit the model

$$Y_{ijk} = \mu + P_i + W_j + (PW)_{ij} + O_{k(ij)}.$$

to both Y = R, the removal rate, and Y = log (U), the log uniformity[5]. This model allows us to evaluate the main effects of P and W, and a possible P by W interaction. As previously stated, P and W are decomposed by orthogonal polynomials, and the component of PW, PQWC, which is confounded with a linear drift, is separated from the rest of PW. The parameter $O_{k(ij)}$ specifies that the wafer-to-wafer variability effects will be computed within each of the 15 experimental conditions.

The results of the analyses of variance are initially shown by half-normal quantile-quantile plots (often called half-normal plots)[6] of the single-degree-of-freedom effects[3], [4]. We discuss the results for removal rate and uniformity separately in Sections 4.3 and 4.4 respectively.

4.3 Results for Removal Rate

Figure 3 presents half-normal plots of single-degree-of-freedom effects from analysis of variance of removal rate with and without the largest effect.

- The platen rotation, P, has by far the most important effect on removal rate and its effect is linear (Figure 3a).

---

5. This is often done when the effects of the design parameters are likely to be multiplicative in the original scale, or because standard deviations are constrained to be non-negative. The log-transformation addresses both of these problems, so that the assumption of a linear model with additive effects, necessary for analysis of variance, is more likely to be satisfied.

6. In the context of analysis of variance, half-normal plots show the absolute values of single-degree-of-freedom effects plotted against quantiles of the standard half-normal distribution. Half-normal rather than full normal plots are used when the magnitude of the effects rather than the sign is of interest. If no parameter has any effect, as in a case of only random variability, the configuration of points in the plot should approximate a straight line. The slope of such a line is an estimate of the standard error, which in this case is the wafer-to-wafer variability. Large effects should appear as points on the right side of the plot that are considerably higher than the other points. When large effects do appear, an estimate of the standard error can be obtained by using the slope after these large effects are removed. Nonhomogeneous variability is indicated when the configuration of points appears to be a set of joined line segments with different slopes.

Figure 3. Half-normal Plots of Single-degree-of-freedom Effects From Analysis of Variance of Removal Rate. Figure 3a shows all effects. P.L, the linear component of P, is much larger than the other effects. Figure 3b shows all effects except P.L. W.L is the linear component of W. PQWC is the linear drift of the polisher.

- The effects of other factors on the removal rate are less clear because of the indication of non-homogeneous variability. (Figure 3b).

Figure 4 shows the effects in Figure 3b separated by type.

- The similar range of Figures 4a and 4b indicates that the wafer-to-wafer variability is as large as the variability due to all parameter effects, except the linear P effect.
- The effect of the polisher drift is as large as the effect of the wafer rotation, W (Figure 4a).
- The wafer-to-wafer variability is not homogeneous (Figure 4b).

To find the cause of the inhomogeneous wafer-to-wafer variability, we plot the variances of the mean removal rates of the three wafers in each P, W combination in Figure 5. The wafer-to-wafer variability of wafers polished with platen rotation P = 19 rpm is clearly larger than that for wafers polished at either P = 11 or 15 rpm. One consequence of the unusual observations for wafer 3 in P = 19 rpm, W = 22 rpm, is seen in the extremely large variance at that P, W combination. Even disregarding this condition, this plot indicates that the higher platen frequency of 19 rpm will not give as good reproducibility as will lower platen frequencies.

Figure 4. Half-normal Plots of Single-degree-of-freedom Effects. Figure 4a shows effects of the parameters omitting P.L, the linear effect of P. W.L, the linear effect of W, and PQWC, the linear effect of the polisher drift, appear larger than the other effects. Figure 4b shows effects of wafer-to-wafer variability. The plotting symbols 1, 2, 3 correspond to order effects within P = 11, 15, and 19 rpm, respectively. It appears that the slope of a line drawn through points labeled with 3 is different from the slope of a line drawn through the remaining points.

We can formally test the inhomogeneity of wafer-to-wafer variability for P = 19 rpm vs P = 11 and 15 rpm by computing an F statistic for the respective pooled variances[5]. Table 2 shows this calculation, with and without the observations at P = 19, W = 22. Clearly the wafer-to-wafer variability is significantly larger for wafers polished at a platen frequency of 19 rpm.

This result presents a problem for analysis of variance. We would like to compare the effects of the parameters to the wafer-to-wafer variability. But one of the assumptions of analysis of variance is homogeneity of the residual variance, which here is the wafer-to-wafer variability. So to quantify the importance of the parameters at the levels in which we are interested, we omit the data for P = 19, and compute an analysis of variance for the remaining data, first correcting for polisher drift. Table 3 presents this analysis of variance. The source W.QC4 is the higher-order components of wafer frequency, and PW is the interaction between the platen and wafer frequencies, now not affected by the drift.

Figure 5. Variances of Mean Removal Rate within Each Polishing Condition. The plotting characters denote levels of P. Wafers polished with P = 19 rpm (plotting character 3) have larger wafer-to-wafer variability.

Table 2. Comparison of Pooled Wafer-to-wafer Variances

Source	DF	Mean Square	F-value	P-value
Variance, P = 19	10	7801.76	8.81	<.0001
Variance, P = 19, omitting W = 22	8	4991.15	5.64	.0008
Variance, P = 11 and 15	20	885.32		

Table 3. Analysis of Variance for Removal Rate Omitting P = 19

Source	DF	Sum of Sq	Mean Sq	F-value	P-value
P	1	246724.6	246724.6	279.31	<.0001
W.L	1	14058.15	14058.15	15.92	.0007
W.QC4	3	384.12	128.04	0.14	0.94
PW	4	2219.8	555.0	0.63	0.65
Wafer-to-wafer	20	17666.3	883.3		

This analysis shows that for platen frequencies 11 and 15:

- The effect of platen frequency, P, is the most important effect in determining removal rate.
- The linear effect of wafer frequency, W.L, is also important.
- There is no non-linear component of wafer frequency.
- There is no interaction between the two rotational frequencies P and W.

There is no quantification of the polisher drift, but we remember from the initial analysis

that the effect of drift was as important as the linear effect of wafer frequency.

We want a model which relates removal rate to the parameters at the lower platen frequencies. The linear relationship between the removal rate and the rotational frequencies, omitting data for P = 19 rpm and correcting for the polisher drift is[7]:

$$\hat{R}_{ij} = \underset{(34.79)}{264.19} + \underset{(2.50)}{45.18}\, P_i + \underset{(0.35)}{1.53}\, W_{j.}$$

This equation quantifies the dependence of removal rate on P and W without being affected by the polisher drift.

Figure 6 shows the results of fitting the model.

Figure 6. Plots of the Fit of a Linear Model in P and W to Removal Rate. Data for P = 19 rpm is omitted. Figure 6a shows the data (uncorrected for drift) and fitted values for P = 11 and 15 rpm. Figure 6b shows a normal quantile-quantile plot of the residuals.

The model appears to fit well, except for the tendency of one or two groups of points to appear a little higher or a little lower than the fit. This is because the fit is corrected for drift, whereas the data are not. For example, the three points at P = 15, W = 22 are higher than the fitted line. These were the first three wafers polished, so the polisher removal

7. Numbers shown in parentheses below the coefficients are the standard errors of the coefficients.

rate was relatively higher. The normal quantile-quantile plot of residuals[8] in Figure 6b has a reasonable linear configuration, indicating that the model is appropriate.

We also evaluated the behavior through time of the residuals from a similar fit including the data for P = 19 rpm. Figure 7 shows a plot of these residuals plotted against order number of each wafer in the experiment. The figure displays the behavior of the polisher through time without the effects due to changing the polishing conditions.

Figure 7. Plot of Residuals from the Fit of Removal Rate to the Linear Effects of P and W. The solid line drawn is the least squares line with slope -1.5 from fitting the plotted residuals to time order. The dashed line has slope 0. The slope of the fitted line is significantly different from 0.

The tendency of the polisher to drift, that is, to show removal rate reduction over time, is seen.

4.4 Results for Uniformity

We defined uniformity within a wafer as the standard deviation about the mean of the amount of oxide removed, actually analyzing the log standard deviation as a function of the design parameters. The analysis was similar to that for removal rate; the results are summarized as follows:

8. Under the assumptions for fitting linear models, the residuals should be normally distributed with mean 0 and some variance. A plot of the ordered residuals vs quantiles from a standard normal distribution should have a straight line configuration with intercept 0 and slope which estimates the residual standard error.

- The linear effects of both platen and wafer frequencies on log uniformity are important.
- There is no P, W interaction effect.
- The uniformity is not affected by the polisher drift.
- There is no indication of nonhomogeneity of wafer-to-wafer variability. In terms of uniformity, the higher platen frequency of 19 rpm seems to give as good results as the lower platen frequencies.

Figure 8a shows the data for uniformity and fitted values computed from the model transformed back to the Angstrom scale, and Figure 8b is a normal plot of the residuals on the log Angstrom scale.

Figure 8. Plots of the Fit of Log Uniformity to P and W. Figure 8a shows the data and fitted values on the Angstrom scale. Figure 8b shows a normal quantile-quantile plot of the residuals on the log Angstrom scale.

The model appears to be reasonable, although the residual variability is large for some polishing conditions. The normal plot shows good linearity in the central part of the distribution, but has tails that are "too-long", indicating that some of the data do not agree with the model as much as we would like. When we investigate the identity of the points in the lower tail of the distribution (better uniformity than expected), we find that all three wafers run at the parameter combination of P = 15 rpm and W = 32 rpm show much better uniformity than is predicted (see Figure 8a). This parameter combination immediately followed the wafer on which there was an interruption due to the polisher

problem noted. The improvement can be due to optimum conditions or to something that changed during the interruption. Since there is no clear trend indicated by the model, an optimum condition is unlikely. The points that have much worse uniformity than predicted (upper tail of Figure 8b) have no consistent values of the experimental parameters.

5. SUMMARY

We have employed a designed experiment to study the effects of platen rotation frequency and wafer rotation frequency on oxide removal rate and uniformity in the chemical-mechanical-polishing process. Some of the important findings are:

- Simple linear models are appropriate for relating both removal rate and uniformity to platen and wafer rotation frequencies.
- Wafer to wafer reproducibility of removal rate is worse for a platen frequency of 19 rpm than for one of 15 or lower rpm. This finding aids in narrowing an operating regime. It also points out that rotational frequencies that result in a more uniform polished oxide also give a lower rate of throughput.
- The effect of polisher drift can be separated from the effects of the rotational frequencies. Removal rate is clearly affected by polisher drift in this experiment; however uniformity is not affected.

This experiment demonstrates that statistical design and analysis are powerful tools in quantifying the various effects that experimenters often observe but may not be able to easily characterize. Besides finding the effects of the design parameters, this experiment helped to quantify the polisher drift that had been previously observed but not well understood, and showed that lower wafer-to-wafer variability is achieved at lower platen frequencies.

Wafer-to-wafer variability is of great concern as it relates to the manufacturability of the chemical-mechanical polishing process. The equipment, operating procedures (such as pad dressing in order to rejuvenate the pad), and process control all could be improved to lower the cost in the routine use of this technology.

6. ACKNOWLEDGEMENTS

G. Smolinsky participated in the planning and execution of this experiment, and in writing a draft of earlier documentation. We also thank a referee who suggested the simplified explanation of determination of the run order shown in the Appendix.

APPENDIX
Details of Run Order

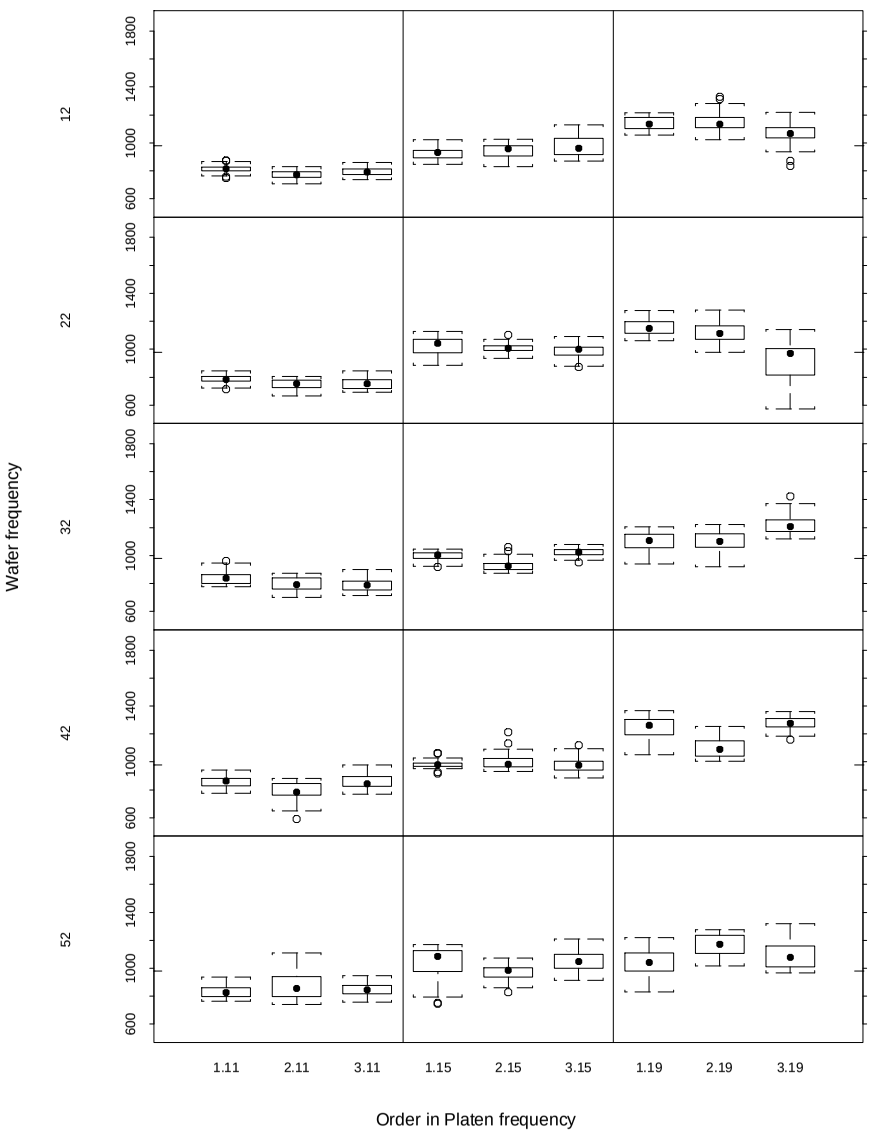
The run order can be determined from the contrast coefficients for the effect to be confounded by ordering the runs in increasing order of these coefficients. In the following table the levels of W and P are shown along with the values of orthogonal polynomial coefficients for the quadratic effect for three levels and the cubic effect for five levels. The body of the table shows the coefficients for the P quadratic by W cubic effect.

		W				
P		12	22	32	42	52
		-1	2	0	-2	1
11	1	-1	2	0	-2	1
15	-2	2	-4	0	4	-2
19	1	-1	2	0	-2	1

A suggested order of runs is then determined by ordering by the PQWC coefficients.

	-4	-2			-1		0			1		2			4
P	15	11	15	19	11	19	11	15	19	11	19	11	15	19	15
W	22	42	52	42	12	12	32	32	32	52	52	22	12	22	42

This ordering lacks randomization, so we randomized within the ties. We also moved one of the runs with W = 32 in a way that would not destroy the confounding, so as not to have all runs with W = 32 together, to get the ordering given in Table 1.


Wafer frequency
12
22
32
42
52
600
1000
1400
1800
1.11
2.11
3.11
1.15
2.15
3.15
1.19
2.19
3.19
Order in Platen frequency

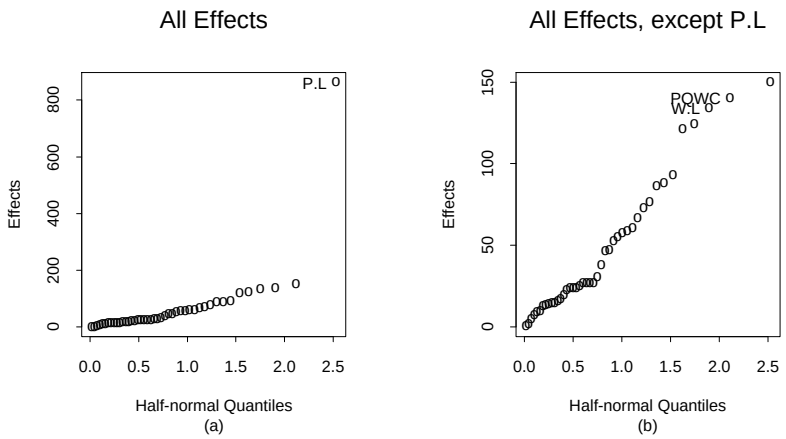

All Effects
P.L
Effects
0
200
400
600
800
0.0
0.5
1.0
1.5
2.0
2.5
Half-normal Quantiles
(a)
All Effects, except P.L
POWC
W:L
Effects
0
50
100
150
0.0
0.5
1.0
1.5
2.0
2.5
Half-normal Quantiles
(b)

FIGURE 3

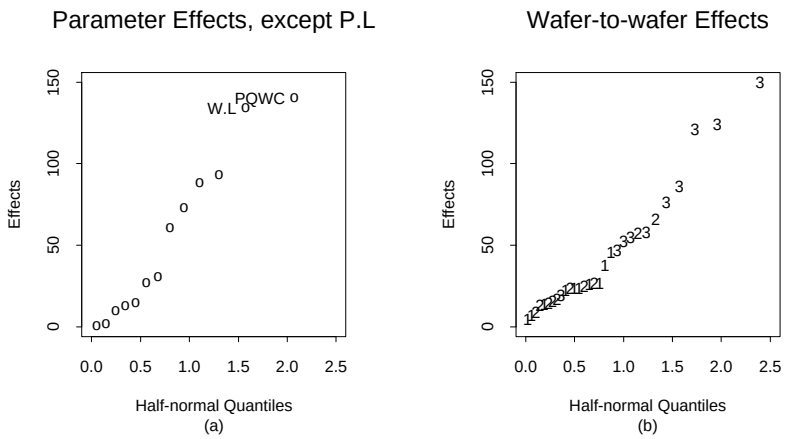

Parameter Effects, except P.L
Effects
0
50
100
150
0.0
0.5
1.0
1.5
2.0
2.5
Half-normal Quantiles
(a)
POWC
W.L
Wafer-to-wafer Effects
Effects
0
50
100
150
0.0
0.5
1.0
1.5
2.0
2.5
Half-normal Quantiles
(b)

FIGURE 4

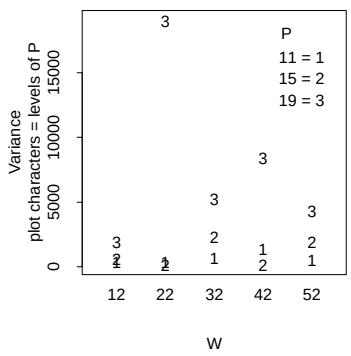

Variance
plot characters = levels of P
0
5000
10000
15000
12
22
32
42
52
W
P
11 = 1
15 = 2
19 = 3

FIGURE 5

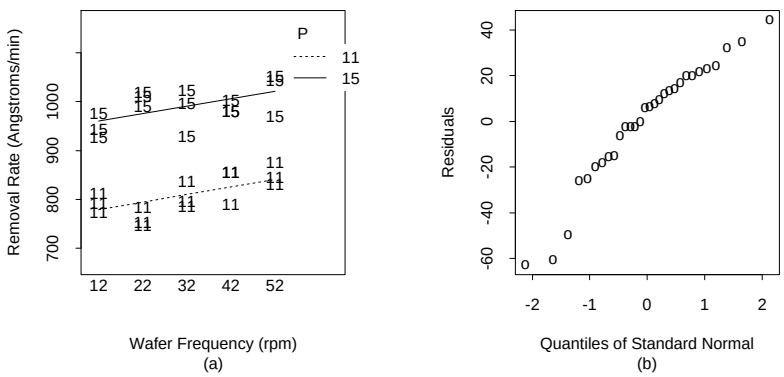

Removal Rate (Angstroms/min)
700
800
900
1000
12
22
32
42
52
Wafer Frequency (rpm)
(a)
P
11
15
Residuals
-60
-40
-20
0
20
40
-2
-1
0
1
2
Quantiles of Standard Normal
(b)

FIGURE 6

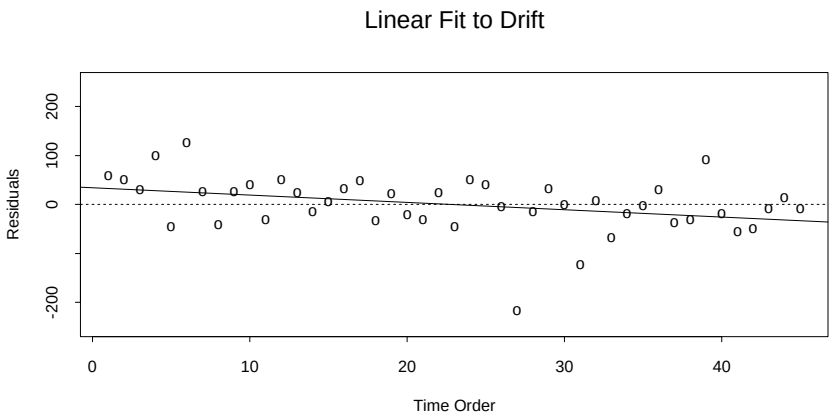

Linear Fit to Drift
Residuals
200
100
0
-200
0
10
20
30
40
Time Order

FIGURE 7

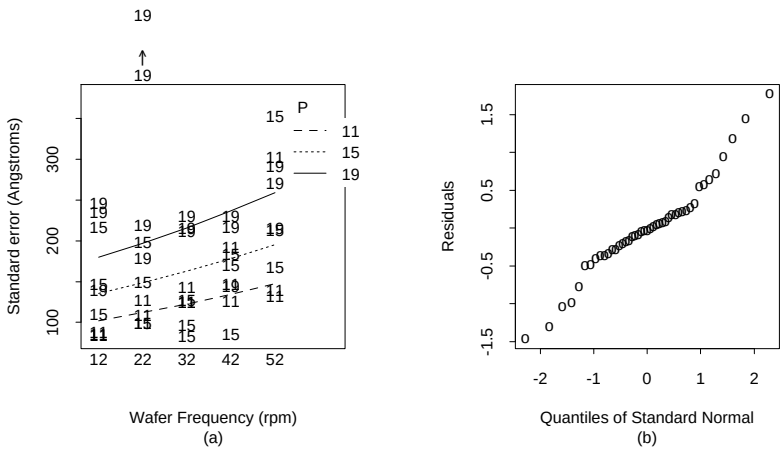

Standard error (Angstroms)
100
200
300
Wafer Frequency (rpm)
12
22
32
42
52
P
11
15
19
(a)
Residuals
-1.5
-0.5
0.5
1.5
Quantiles of Standard Normal
-2
-1
0
1
2
(b)

FIGURE 8

REFERENCES

1. N. Lifshitz, A. J. Miller, W. Y-C. Lai, G. Smolinsky, Planarization of the Interlevel Dielectrics by Polishing: Demonstration of Feasibility, AT&T Bell Laboratories Internal Memorandum.

2. C. R. Hicks, *Fundamental Concepts in the Design of Experiments*, Holt, Rinehart and Winston, New York, 1973, pp. 131-151.

4. C. Daniel, "Use of Half-normal Plots in Interpreting Factorial Two-level Experiments", *Technometrics*, Vol 1, No. 4, November 1959.

4. V. N. Nair and D. Pregibon, "A Data Analysis Strategy for Quality Engineering Experiments", AT&T Technical Journal, Vol 65, Issue 3, May/June 1986, pp. 73-84.

5. J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, Wadsworth, Brooks/Cole, Monterey, CA, 1982, pp. 311-313.