Copyright © 1983 American Telephone and Telegraph Company
THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 62, No. 5, May–June 1983
Printed in U.S.A.

Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design

By M. S. PHADKE, R. N. KACKAR, D. V. SPEENEY, and M. J. GRIECO

(Manuscript received March 25, 1982)

In this paper we describe the off-line quality control method and its application in optimizing the process for forming contact windows in 3.5-μm complementary metal-oxide semiconductor circuits. The off-line quality control method is a systematic method of optimizing production processes and product designs. It is widely used in Japan to produce high-quality products at low cost. The key steps of off-line quality control are: (i) Identify important process factors that can be manipulated and their potential working levels; (ii) perform fractional factorial experiments on the process using orthogonal array designs; (iii) analyze the resulting data to determine the optimum operating levels of the factors (both the process mean and the process variance are considered in this analysis; (iv) conduct an additional experiment to verify that the new factor levels indeed improve the quality control.

I. INTRODUCTION AND SUMMARY

This paper describes and illustrates the off-line quality control method, which is a systematic method of optimizing a production process. It also documents our efforts to optimize the process for forming contact windows in 3.5-μm technology complementary metal-oxide semiconductor (CMOS) circuits fabricated in the Murray Hill Integrated Circuit Design Capability Laboratory (MH ICDCL). Here, by optimization we mean minimizing the process variance while keeping the process mean on target.

A typical very large scale integrated circuit (IC) chip has thousands of contact windows (e.g., a *BELLMAC**-32 microprocessor chip has

* Trademark of Bell Laboratories.

250,000 windows on an approximately 1.5-cm^2 area), most of which are not redundant. It is critically important to produce windows of size very near the target dimension. (In this paper windows mean contact windows.) Windows that are not open or are too small result in loss of contact to the devices, while excessively large windows lead to shorted device features. The application of the off-line quality control method has reduced the variance of the window size by a factor of four. Also, it has substantially reduced the processing time required for the window-forming step.

This study was inspired by Professor Genichi Taguchi's visit to the Quality Theory and Systems Group in the Quality Assurance Center at Bell Laboratories during the months of August, September, and October, 1980. Professor Taguchi, director of the Japanese Academy of Quality and a recipient of the Deming award, has developed the method of off-line quality control during the last three decades. It is used routinely by many leading Japanese industries to produce high-quality products at low cost. An overview of Professor Taguchi's off-line and on-line quality control methods is given in Taguchi,[1] and Kackar and Phadke.[2] This paper documents the results of the first application of Professor Taguchi's off-line quality control method in Bell Laboratories.

The distinctive features of the off-line quality control method are experimental design using orthogonal arrays and the analysis of signal-to-noise ratios (s/n). The orthogonal array designs provide an economical way of simultaneously studying the effects of many production factors on the process mean and variance. Orthogonal array designs are fractional factorial designs with the orthogonality property defined in Section IV. The s/n is a measure of the process variability. According to Professor Taguchi,[3] by optimizing the process with respect to the s/n, we ensure that the resulting optimum process conditions are robust or stable, meaning that they have the minimum process variation.

The outline of this paper is as follows: Section II gives a brief description of the window-forming process, which is a critical step in IC fabrication. The window-forming process is generally considered to be one of the most difficult steps in terms of reproducing and obtaining uniform-size windows. Nine key process factors were identified and their potential operating levels were determined. A description of the factors and their levels is given in Section III. The total number of possible factor-level combinations is about six thousand.

The aim of the off-line quality control method is to determine a factor-level combination that gives the least variance for the window size while keeping the mean on target. To determine such a factor-level combination we performed eighteen experiments using the L_{18}

orthogonal array. The experimental setup is given in Section IV. These eighteen experiments correspond to eighteen factor-level combinations among the possible six thousand combinations. For each experiment, measurements were taken on the line width and the window-size control features. The resulting data were analyzed to determine the optimum factor-level combination. The measurements and the data analysis are presented in Sections V through IX.

The optimum factor levels, inferred from the data analysis, were subsequently used in fabricating the *BELLMAC*-32 microprocessor, the *BELLMAC*-4 microcomputer, and some other chips in the Murray Hill ICDCL. The experience of using these conditions is discussed in Section X.

The experiment was designed and preliminary analysis of the experimental data was performed under Professor Taguchi's guidance and collaboration.

II. THE WINDOW-FORMING PROCESS

Fabrication of integrated circuits is a complex, lengthy process.[4] Window forming is one of the more critical steps in fabricating state of the art CMOS integrated circuits. It comes after field and gate oxides are grown; polysilicon lines have been formed; and the gate, source, and drain areas are defined by the process of doping. Figure 1 shows the windows in a cross section of a wafer. A window is a hole of about 3.5 μm diameter etched through an oxide layer of about 2 μm thickness. The purpose of the windows is to facilitate the interconnections between the gates, sources, and drains. For this reason these windows are called contact windows.

The process of forming windows through the oxide layers involves photolithography. First the *P*-glass surface is prepared by depositing

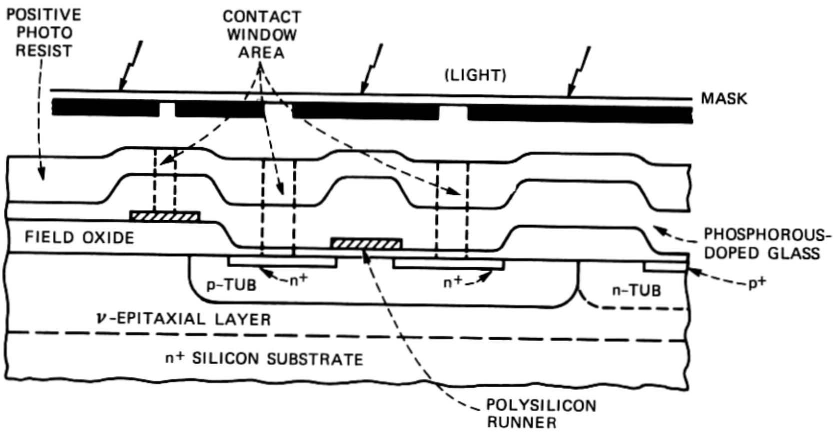

Fig. 1—Cross section of a wafer.

undoped oxide on it and prebaking it. The window-forming process is described below.

(*i*) Apply Photoresist: A wetting agent is sprayed on the wafer to promote adhesion of photoresist to the oxide surface. Then an appropriate photoresist is applied on the wafer and the wafer is rotated at high speed so that the photoresist spreads uniformly.

(*ii*) Bake: The wafer is baked to dry the photoresist layer. The thickness of the photoresist layer at this stage is about 1.3 to 1.4 μm.

(*iii*) Expose: The photoresist-coated wafer is exposed to ultraviolet radiation through a mask. The windows to be printed appear as clear areas on the mask. In addition to the windows, which are parts of the desired circuits, the mask has some test patterns. Light passes through these areas and causes the photoresist in the window areas and the test pattern areas to become soluble in an appropriate solvent (developer). The areas of the photoresist where light does not strike remain insoluble.

(*iv*) Develop: The exposed wafer is dipped in the developer, which dissolves only the exposed areas. In properly printed windows, the exposed photoresist is removed completely and the oxide surface is revealed.

(*v*) Plasma Etch: The wafers are placed in a high-vacuum chamber wherein a plasma is established. The plasma etches the exposed oxide areas faster than it etches the photoresist. So at the places where the windows are printed, windows are cut through the oxide layers down to the silicon surface.

(*vi*) Remove Photoresist: The remaining photoresist is now removed with the help of oxygen plasma and wet chemicals.

In the formation of the final contact windows there are additional steps: (*vii*) removal of cap-oxide, (*viii*) oxidation of the contact area to prevent diffusion of phosphorus in the subsequent step, (*ix*) reflow of the *P*-glass to round the window corners, (*x*) hydrogen annealing, and (*xi*) pre-metal wet-etching to remove any remaining oxides from the contact window areas.

At the time we started this study, the target window size at step 6 was considered to be 3.0 μm. The final target window size (after step *xi*) was 3.5 μm.

III. SELECTION OF FACTORS AND FACTOR LEVELS

For the present study only the steps numbered (*i*) through (*v*) were chosen for optimization. Discussions with process engineers led to the selection of the following nine factors for controlling the window size. The factors are shown next to the appropriate fabrication steps.

(*i*) Apply Photoresist: Photoresist viscosity (B) and spin speed (C).

(*ii*) Bake: Bake temperature (D) and bake time (E).

(*iii*) Expose: Mask dimension (A), aperture (F), and exposure time (G).

(*iv*) Develop: Developing time (H).

(*v*) Plasma etch: Etch time (I).

No factor was chosen corresponding to the photoresist removal step because it does not affect the window size.

The standard operating levels of the nine factors are given in Table I. Under these conditions, which prevailed in September 1980, the contact windows varied substantially in size and on many occasions even failed to print and open. Figure 2 shows a typical photograph of the programmed logic array (PLA) area of a microcomputer chip. The wide variation in window size and the presence of unopened windows is obvious from the figure.

The principle of off-line quality control is to systematically investigate various possible levels for these factors with an aim of obtaining uniform-size windows.

In the window-forming experiment a number of alternate levels were considered for each of the nine factors. These levels are also listed in Table I. Six of these factors have three levels each. Three of the factors have only two levels.

Table I—Test levels

Label	Factors Name	Levels		
			Standard Levels	
A	Mask Dimension (μm)		2	2.5
B	Viscosity		204	206
C	Spin Speed (rpm)	Low	Normal	High
D	Bake Temperature (°C)	90	105	
E	Bake Time (min)	20	30	40
F	Aperture	1	2	3
G	Exposure Time	20% Over	Normal	20% Under
H	Developing Time (s)	30	45	60
I	Plasma Etch Time (min)	14.5	13.2	15.8

Dependence of spin speed on viscosity

		Spin Speed (rpm)		
		Low	Normal	High
Viscosity	204	2000	3000	4000
	206	3000	4000	5000

Dependence of exposure on aperture

		Exposure (PEP-Setting)		
		20% Over	Normal	20% Under
Aperture	1	96	120	144
	2	72	90	108
	3	40	50	60

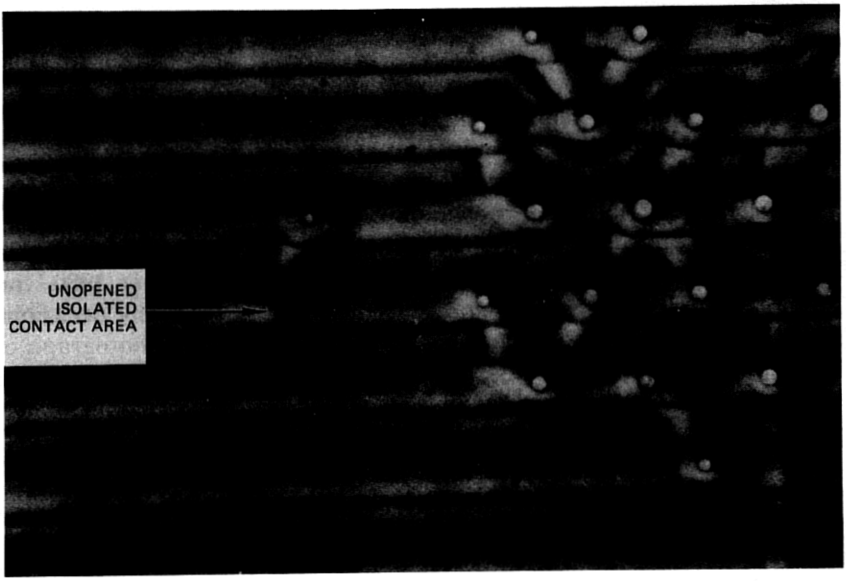

Fig. 2—Example of nonuniform contact window sizes and an isolated, unopened contact window. Both are typical results obtained in August 1980, for the PLA area of microprocessor and microcomputer chips. The contact windows are round shaped.

The levels of spin speed are tied to the levels of viscosity. For the 204 photoresist viscosity the low, normal, and high spin speeds mean 2000 rpm, 3000 rpm, and 4000 rpm, respectively. For the 206 photoresist viscosity the spin speed levels are 3000 rpm, 4000 rpm, and 5000 rpm. Likewise, the exposure setting depends on the aperture. These relationships are also shown in Table I.

IV. THE ORTHOGONAL ARRAY EXPERIMENT

The full factorial experiment to explore all possible factor-level combinations would require $3^6 \times 2^3 = 5832$ experiments. Considering the cost of material, the time, and the availability of facilities, the full factorial experiment is prohibitively large. Also from statistical considerations it is unnecessary to perform the full factorial experiment because processes can usually be adequately characterized by a relatively few parameters.

The fractional factorial design used for this study is given in Table II. It is the L_{18} orthogonal array design consisting of 18 experiments taken from Taguchi and Wu.[3] The rows of the array represent runs while the columns represent the factors. Here we treat BD as a joint factor with the levels 1, 2, and 3 representing the combinations B_1D_1, B_2D_1, and B_1D_2, respectively. This is done so that we can study all the

Table II—The L_{18} orthogonal array

Experiment Number	1 A	2 BD	3 C	4 E	5 F	6 G	7 H	8 I
	Column Number & Factor							
1	1	1	1	1	1	1	1	1
2	1	1	2	2	2	2	2	2
3	1	1	3	3	3	3	3	3
4	1	2	1	1	2	2	3	3
5	1	2	2	2	3	3	1	1
6	1	2	3	3	1	1	2	2
7	1	3	1	2	1	3	2	3
8	1	3	2	3	2	1	3	1
9	1	3	3	1	3	2	1	2
10	2	1	1	3	3	2	2	1
11	2	1	2	1	1	3	3	2
12	2	1	3	2	2	1	1	3
13	2	2	1	2	3	1	3	2
14	2	2	2	3	1	2	1	3
15	2	2	3	1	2	3	2	1
16	2	3	1	3	2	3	1	2
17	2	3	2	1	3	1	2	3
18	2	3	3	2	1	2	3	1

nine factors with the L_{18} orthogonal array. Thus, experiment 2 would be run under level 1 of factors A, B, and D, and level 2 of the remaining factors. In terms of the actual settings, these conditions are: 2-μm mask dimension, 204 viscosity, 90°C bake temperature, 3000-rpm spin speed, bake time of 30 minutes, aperture 2, exposure PEP setting 90, 45-second developing time, and 13.2 minutes of plasma etch. The other rows are interpreted similarly.

Here are some of the properties and considerations of this design:

(*i*) This is a main-effects-only design; i.e., the response is approximated by a separable function. A function of many independent variables is called separable if it can be written as a sum of functions where each component function is a function of only one independent variable.

(*ii*) For estimating the main effects there are two degrees of freedom associated with each three-level factor, one degree of freedom for each two-level factor, and one degree of freedom with the overall mean. We need at least one experiment for every degree of freedom. Thus, the minimum number of experiments needed is $2 \times 6 + 1 \times 3 + 1 = 16$. Our design has 18 experiments. A single-factor-by-single-factor experiment would need only 16 experiments, two fewer than 18. But such an experiment would yield far less precise information compared with the orthogonal array experiment.[3,5]

(*iii*) The columns of the array are pairwise orthogonal. That is, in every pair of columns, all combinations of levels occur and they occur an equal number of times.

(*iv*) Consequently, the estimates of the main effects of all factors as shown in Table II and their associated sums of squares are independent under the assumption of normality and equality of error variance. So the significance tests for these factors are independent. Though *BD* is treated as a joint factor, the main effects and sums of squares of *B* and *D* can be estimated separately under the assumption of no interaction. In general, these estimates would be correlated with each other. However, these estimates are not correlated with those for any of the other seven factors.

(*v*) The estimates of the main effects can be used to predict the response for any combination of the parameter levels. A desirable feature of this design is that the variance of the prediction error is the same for all parameter-level combinations covered by the full factorial design.

(*vi*) It is known that the main-effect-only models are liable to give misleading conclusions in the presence of interactions. However, in the beginning stages of this study the interactions are assumed to be negligible. If we wished to study all two-factor interactions, with no more than 18 experiments we would have enough degrees of freedom for studying only two three-level factors, or five two-level factors! That would mean in the present study we would have to eliminate half of the process factors without any experimental evidence. Alternately, if we wished to study all the nine process factors and their two-factor interactions, we would need at least 109 experiments! Orthogonal array designs can, of course, be used to study interactions.[3]

(*vii*) Optimum conditions obtained from such an experiment have to be verified with an additional experiment. This is done to safeguard us against the potential adverse effects of ignoring the interactions among the manipulatable factors.

In conducting experiments of this kind, it is common for some wafers to get damaged or broken. Also, the wafer-to-wafer variability of window sizes is typically large. So we decided to run each experiment with two wafers.

4.1 Analysis of variance

Data collected from such experiments are analyzed by a method called analysis of variance (ANOVA).[6] The purpose of ANOVA is to separate the total variability of the data, which is measured by the sum of the squared deviations from the mean value, into contributions by each of the factors and the error. This is analogous to the use of Parseval's theorem to separate the signal strength into contributions by the various harmonics.[3] To see which of the factors have a significant effect, *F*-tests are performed. In performing the standard *F*-test we assume that the errors are normally distributed with equal variance and are independent. The results of the *F*-test are indicated by the significance level. When we say that a factor is significant at 5-percent level we mean that there is 5 percent or less chance that, if we change the level of the factor, the response will remain the same. If the F-test indicates that a factor is not significant at the 5-percent level it means that, if we change the level of that factor, there is more than a 5-percent chance that the response will remain the same.

The levels of factors which are identified as significant are then set to obtain the best response. The levels of the other factors can be set at any levels within the experimental range. We choose to leave them at the starting levels.

If the assumptions of the F-test are not completely satisfied, the quoted significances are not accurate. However, the standard F-test is relatively insensitive to deviations from the assumptions used in its derivation. Thus, for making engineering decisions about which factor levels to change, the accuracy of the significance level is an adequate guide. In this paper we will use the standard F-test even though some of the assumptions are not strictly satisfied.

V. QUALITY MEASURES

The window size is the relevant quality measure for this experiment. The existing equipment does not give reproducible measurements of the sizes of windows in the functional circuits on a chip. This is because of the small size of these windows and their close proximity to one another. Therefore, test patterns—a line-width pattern and a window pattern—are provided in the upper left-hand corner of each chip. The following measurements were made on these test patterns to indicate the quality.

(*i*) Line width after step (*iv*), called the pre-etch line width or photo-line width.

(*ii*) Line width after step (*vi*), called the post-etch line width.

(*iii*) Size of the window test pattern after step (*vi*), called the post-etch window size.

Five chips were selected from each wafer for making the above measurements. These chips correspond to specific locations on a wafer—top, bottom, left, right, and center.

All three quality measures are considered to be good indicators of the size of the functional windows. However, between the geometries of the window-size pattern and the line-width pattern, the geometry of the window-size pattern is closer to the geometry of the functional windows. So, among the three quality measures, the post-etch window size may be expected to be better correlated with the size of the functional windows.

VI. EXPERIMENTAL DATA

Only thirty-four wafers were available for experimentation. So experiments 15 and 18 were arbitrarily assigned only one wafer each. One

Table III—Experimental data

Experiment No.	Line-Width Control Feature Photoresist—Nanoline Tool (Micrometers)					Comments
	Top	Center	Bottom	Left	Right	
1	2.43	2.52	2.63	2.52	2.5	
1	2.36	2.5	2.62	2.43	2.49	
2	2.76	2.66	2.74	2.6	2.53	
2	2.66	2.73	2.95	2.57	2.64	
3	2.82	2.71	2.78	2.55	2.36	
3	2.76	2.67	2.9	2.62	2.43	
4	2.02	2.06	2.21	1.98	2.13	
4	1.85	1.66	2.07	1.81	1.83	
5	—	—	—	—	—	Wafer Broke
5	1.87	1.78	2.07	1.8	1.83	
6	2.51	2.56	2.55	2.45	2.53	
6	2.68	2.6	2.85	2.55	2.56	
7	1.99	1.99	2.11	1.99	2.0	
7	1.96	2.2	2.04	2.01	2.03	
8	3.15	3.44	3.67	3.09	3.06	
8	3.27	3.29	3.49	3.02	3.19	
9	3.0	2.91	3.07	2.66	2.74	
9	2.73	2.79	3.0	2.69	2.7	
10	2.69	2.5	2.51	2.46	2.4	
10	2.75	2.73	2.75	2.78	3.03	
11	3.2	3.19	3.32	3.2	3.15	
11	3.07	3.14	3.14	3.13	3.12	
12	3.21	3.32	3.33	3.23	3.10	
12	3.48	3.44	3.49	3.25	3.38	
13	2.6	2.56	2.62	2.55	2.56	
13	2.53	2.49	2.79	2.5	2.56	
14	2.18	2.2	2.45	2.22	2.32	
14	2.33	2.2	2.41	2.37	2.38	
15	2.45	2.50	2.51	2.43	2.43	
15	—	—	—	—	—	No wafer
16	2.67	2.53	2.72	2.7	2.6	
16	2.76	2.67	2.73	2.69	2.6	
17	3.31	3.3	3.44	3.12	3.14	
17	3.12	2.97	3.18	3.03	2.95	
18	3.46	3.49	3.5	3.45	3.57	
18	—	—	—	—	—	No wafer

of the wafers assigned to experiment 5 broke in handling. So experiments 5, 15, and 18 have only one wafer.

The experimental data are shown in Table III.

The data arising from such experiments can be classified as two types—continuous data and categorical data. Here, the pre-etch and the post-etch line-width data are of the continuous type. The post-etch window size data are mixed categorical-continuous type, because some windows are open while some are not. The two types of data are analyzed somewhat differently, as we explain the following two sections.

VII. ANALYSIS OF THE LINE-WIDTH DATA

Both the pre-etch and the post-etch line widths are continuous variables. For each of these variables the statistics of interest are the

Table III—Experimental data (Continued)

Experiment No.	Line-Width Control Feature Etched—Nanoline Tool (Micrometers)					Comments
	Top	Center	Bottom	Left	Right	
1	2.95	2.74	2.85	2.76	2.7	
1	3.03	2.95	2.75	2.82	2.85	
2	3.05	3.18	3.2	3.16	3.06	
2	3.25	3.15	3.09	3.11	3.16	
3	3.69	3.57	3.78	3.55	3.40	
3	3.92	3.62	3.71	3.71	3.53	
4	2.68	2.62	2.9	2.45	2.7	
4	2.29	2.31	2.77	2.46	2.49	
5	—	—	—	—	—	Wafer Broke
5	1.75	1.15	2.07	2.12	1.53	
6	3.42	2.98	3.22	3.13	3.17	
6	3.34	3.21	3.23	3.25	3.28	
7	2.62	2.49	2.53	2.41	2.51	
7	2.76	2.94	2.68	2.62	2.51	
8	4.13	4.38	4.41	4.03	4.03	
8	4.0	4.02	4.18	3.92	3.91	
9	3.94	3.82	3.84	3.57	3.71	
9	3.44	3.30	3.41	3.28	3.20	
10	3.17	2.85	2.84	3.06	2.94	
10	3.70	3.34	3.45	3.41	3.29	
11	4.01	3.91	3.92	3.80	3.90	
11	3.67	3.31	2.86	3.41	3.23	
12	4.04	3.80	4.08	3.81	3.94	
12	4.51	4.37	4.45	4.24	4.48	
13	3.40	3.12	3.11	3.25	3.06	
13	3.22	3.03	2.89	2.92	2.98	
14	3.18	3.03	3.4	3.17	3.32	
14	3.18	2.83	3.17	3.07	3.02	
15	2.86	2.46	2.3	2.6	2.55	
15	—	—	—	—	—	No wafer
16	2.85	2.14	1.22	2.8	3.03	
16	3.4	2.97	2.96	2.87	2.88	
17	4.06	3.87	3.90	3.94	3.87	
17	4.02	3.49	3.51	3.69	3.47	
18	4.49	4.28	4.34	4.39	4.25	
18	—	—	—	—	—	No wafer

mean and the standard deviation. The objective of our data analysis is to determine the factor-level combination such that the standard deviation is minimum while keeping the mean on target. We will call this the optimum factor-level combination. Professor Taguchi's method for obtaining the optimum combination is given next.

7.1 Single response variable

Let us first consider the case where there is only one response variable. Instead of working with the mean and the standard deviation, it is preferable to work with the transformed variables—the mean and the signal-to-noise ratio (s/n). The s/n is defined as

$$s/n = \log_{10}\left(\frac{\text{Mean}}{\text{Standard Deviation}}\right)$$

$$= -\log_{10}(\text{coefficient of variation}).$$

Table III—Experimental data (Continued)

Experiment No.	Window-Control Feature Etched—Vickers Tool (Micrometers) Top	Center	Bottom	Left	Right	Comments
1	WNO*	WNO	WNO	WNO	WNO	
1	WNO	WNO	WNO	WNO	WNO	
2	2.32	2.23	2.30	2.56	2.51	
2	2.22	2.33	2.34	2.15	2.35	
3	2.98	3.14	3.02	2.89	3.16	
3	3.15	3.08	2.78	WNO	2.86	
4	WNO	WNO	WNO	WNO	WNO	
4	WNO	WNO	WNO	WNO	WNO	
5	—	—	—	—	—	Wafer Broke
5	WNO	WNO	WNO	WNO	WNO	
6	2.45	2.19	2.14	2.32	2.12	
6	WNO	WNO	WNO	WNO	WNO	
7	WNO	WNO	WNO	WNO	WNO	
7	WNO	WNO	WNO	WNO	WNO	
8	WNO	WNO	WNO	WNO	WNO	
8	2.89	2.97	3.13	3.25	3.19	
9	3.16	2.91	3.12	3.18	3.11	
9	2.43	2.35	2.14	2.40	2.28	
10	2.0	1.75	1.97	1.91	1.72	
10	WNO	2.7	WNO	2.61	2.73	
11	2.76	3.09	3.22	3.05	3.04	
11	3.12	3.21	WNO	2.71	2.27	
12	3.24	3.08	WNO	2.89	2.72	
12	3.5	3.71	3.52	3.53	3.71	
13	2.54	2.63	2.88	2.31	2.71	
13	WNO	WNO	WNO	WNO	WNO	
14	WNO	1.74	2.24	2.07	2.38	
14	WNO	WNO	WNO	WNO	WNO	
15	WNO	WNO	WNO	WNO	WNO	
15	—	—	—	—	—	No wafer
16	WNO	WNO	WNO	WNO	WNO	
16	WNO	WNO	WNO	WNO	WNO	
17	3.09	2.91	3.06	3.09	3.29	
17	3.39	2.5	2.57	2.62	2.35	
18	3.39	3.34	3.45	3.44	3.33	
18	—	—	—	—	—	No wafer

* WNO—Window not open.

In terms of the transformed variables, the optimization problem is to determine the optimum factor levels such that the s/n is maximum while keeping the mean on target. This problem can be solved in two stages:

(*i*) Determine which factors have a significant effect on the s/n. This is done through the analysis of variance (ANOVA) of the s/n. These factors are called the *control factors*, implying that they control the process variability. For each control factor we choose the level with the highest s/n as the optimum level. Thus the overall s/n is maximized.

(*ii*) Select a factor that has the smallest effect on the s/n among all factors that have a significant effect on the mean. Such a factor is called a *signal factor*. Ideally, the signal factor should have no effect on the s/n. Choose the levels of the remaining factors (factors that are neither control factors nor signal factors) to be the nominal levels prior to the optimization experiment. Then set the level of the signal factor so that the mean response is on target.

In practice, the following two aspects should also be considered in selecting the signal factor: (*i*) If possible, the relationship between the mean response and the levels of the signal factor should be linear, and (*ii*) It should be convenient to change the signal factor during production. These aspects are important from the on-line quality control considerations. The signal factor can be used during manufacturing to adjust the mean response.[1–3]

Why do we work in terms of the s/n ratio rather than the standard deviation? Frequently, as the mean decreases, the standard deviation also decreases and vice versa. In such cases, if we work in terms of the standard deviation, the optimization cannot be done in two steps; i.e., we cannot minimize the standard deviation first and then bring the mean on target.

Through many applications, Professor Taguchi has empirically found that the two-stage optimization procedure involving the s/n indeed gives the parameter-level combination where the standard deviation is minimum, while keeping the mean on target. This implies that the engineering systems behave in such a way that the manipulatable production factors can be divided into three categories:

(*i*) Control factors, which affect process variability as measured by the s/n

(*ii*) Signal factors, which do not influence (or have negligible effect on) the s/n but have a significant effect on the mean

(*iii*) Factors that do not affect the s/n or the process mean.

The two-stage procedure also has an advantage over a procedure that directly minimizes the mean square error from the target mean value. In practice, the target mean value may change during the process development. The advantage of the two-stage procedure is that for any target mean value (of course, within reasonable bounds) the new optimum factor-level combination is obtained by suitably adjusting the level of only the signal factor. This is so because in step (*i*) of the algorithm the coefficient of variation is minimized for every mean target value.

7.2 Multiple response variables

Now let us consider the case where there are two or more response variables. In such cases, engineering judgment may have to be used to resolve the conflict if different response variables suggest different levels for any one factor. The modified two-stage procedure is as follows:

(*i*) Separately determine control factors and their optimum levels corresponding to each response variable. If there is a conflict between the optimum levels suggested by the different response variables, use engineering judgment to resolve the conflict.

(*ii*) Select a factor that has the smallest effect (preferably no effect) on the signal-to-noise ratios for all the response variables but has a significant effect on the mean levels. This is the signal factor. Set the levels of the remaining factors, which affect neither the mean nor the s/n, at the nominal levels prior to the optimization experiment. Then set the level of the signal factor so that the mean responses are on target. Once again engineering judgment may have to be used to resolve any conflicts that arise.

The selection of the control factors, signal factor, and their optimum levels for the present application will be discussed in Section IX. The remaining portions of Sections VII and VIII contain the data analysis that forms the basis for selecting the optimum factor levels.

7.3 Pre-etch line width

Mean, standard deviation, and s/n were calculated for each of the eighteen experiments. For those experiments with two wafers, ten data points were used in these calculations. When there was only one wafer, five data points were used. These results are shown in Table IV. The presence of unequal sample sizes has been ignored in the subsequent analysis. Let $\bar{x}_i$ and η_i denote the mean and the s/n for the ith experiment.

Table IV—Pre-etch line-width data

Experiment Number	Mean Line Width, $\bar{x}$ (μm)	Standard Deviation of Line Width, s (μm)	s/n $\eta = \log(\bar{x}/s)$
1	2.500	0.0827	1.4803
2	2.684	0.1196	1.3512
3	2.660	0.1722	1.1889
4	1.962	0.1696	1.0632
5	1.870	0.1168	1.2043
6	2.584	0.1106	1.3686
7	2.032	0.0718	1.4520
8	3.267	0.2101	1.1917
9	2.829	0.1516	1.2709
10	2.660	0.1912	1.1434
11	3.166	0.0674	1.6721
12	3.323	0.1274	1.4165
13	2.576	0.0850	1.4815
14	2.308	0.0964	1.3788
15	2.464	0.0385	1.8065
16	2.667	0.0706	1.5775
17	3.156	0.1569	1.3036
18	3.494	0.0473	1.8692

By computing a single mean $\bar{x}_i$ and a single variance s_i^2 (needed for computing η_i) from the two wafers of each experiment i, we pool together the between wafer and the within wafer variance. That is,

$$E(s_i^2) = (\text{Between wafer variance for experiment } i) \times 5/9$$
$$+ (\text{Within wafer variance for experiment } i).$$

Thus, when we maximize η, we minimize the sum of the between-wafer and the within-wafer variances of the line width, which is the response of interest to us. There can be situations when one wants to separately estimate the effects of the factor levels on the between-wafer and within-wafer variances. In those cases, one would compute the s/n and the mean line width for each individual wafer.

In the analysis of the pre-etch and the post-etch line widths, we compute the $\bar{x}_i$ and the s_i^2 for each experiment by pooling the data from both wafers used in that experiment. A relative measure of the between-wafer and within-wafer variance is obtained in Section VIII, while the post-etch window-size data is being analyzed.

7.3.1 Analysis of s/n

The estimates of the average s/n for all factor levels are given in Table V. The average for the first level of factor A is the average of the nine experiments (experiments 1 through 9), which were conducted with level 1 of the factor A. Likewise, the average for the second level of factor A is the mean of experiments 10 through 18, which were conducted with level 2 of the factor A. Let us denote these average effects of A_1 and A_2 by m_{A_1} and m_{A_2}, respectively. Here $m_{A_1} = 1.2857$ and $m_{A_2} = 1.5166$. The other entries of Table V were calculated similarly.

The average signal-to-noise ratios for every level of the eight factors are graphically shown in Fig. 3. Qualitatively speaking, the mask dimension and the aperture cause a large variation in the s/n. The

Table V—Pre-etch line width for average s/n

	Factor	Average s/n		
		Level 1	Level 2	Level 3
A	Mask Dimension	1.2857	1.5166	
BD	Viscosity Bake Temperature	(B_1D_1) 1.3754	(B_2D_1) 1.3838	(B_1D_2) 1.4442
B	Viscosity	1.4098	1.3838	
D	Bake Temperature	1.3796	1.4442	
C	Spin Speed	1.3663	1.3503	1.4868
E	Bake Time	1.4328	1.4625	1.3082
F	Aperture	1.5368	1.4011	1.2654
G	Exposure Time	1.3737	1.3461	1.4836
H	Developing Time	1.3881	1.4042	1.4111

Overall average s/n = 1.4011.

developing time and the viscosity cause a small change in the s/n. The effect of the other factors is in between.

For three-level factors, Fig. 3 can also be used to judge the linearity of the effect of the factors. If the difference between levels 1 and 2, and levels 2 and 3 is equal and these levels appear in proper order (1, 2, 3, or 3, 2, 1), then the effect of that factor is linear. If either the differences are unequal or the order is mixed up, then the effect is not linear. For example, the aperture has approximately linear response while the bake time has a nonlinear response.

We shall perform a formal analysis of variance (ANOVA) to identify statistically significant factors. The analysis of variance of general linear models is widely known in literature, see e.g., Searle[7] and Hicks.[6] Simple ANOVA methods for orthogonal array experiments are described in Taguchi and Wu.[3] The linear model used in analyzing this data is:

$$y_i = \mu + x_i + e_i, \tag{1}$$

where

$i = 1, \cdots, 18$ is the experiment number.

μ is the overall mean.

x_i is the fixed effect of the factor-level combination used in experiment i. Here we consider only the main effect for each of the factors. Thus it represents the sum of the effects of the eight factors.

e_i is the random error for experiment i.

y_i is the s/n for experiment i.

To clarify the meaning of the term x_i, let us consider experiment 1, which was run at level 1 of each of the eight factors A through H. Note that the factor I is irrelevant for studying the pre-etch line width. So x_1 is the sum of the main effects associated with the first level of each of the factors A through H.

The sum of squares and the mean squares for the eight factors are tabulated in Table VIa. The computations are illustrated in Appendix A.

The expected mean squares are also shown in Table VIa. See Refs. 6 and 7 for the computation of expected mean squares, which are used in forming appropriate F-tests. The error variance, i.e., variance of e_i, is denoted by σ^2. The variability due to the factors A through H is denoted by ϕ with an appropriate subscript.

In Table VIa we see that the mean sum of squares for factors BD, C, G, and H are smaller than the mean error sum of squares. So a new ANOVA table, Table VIb, was formed by pooling the sum of squares of these factors with the error sum of squares. The linear model underlying the ANOVA Table VIb is the same as Eq. (1), except that

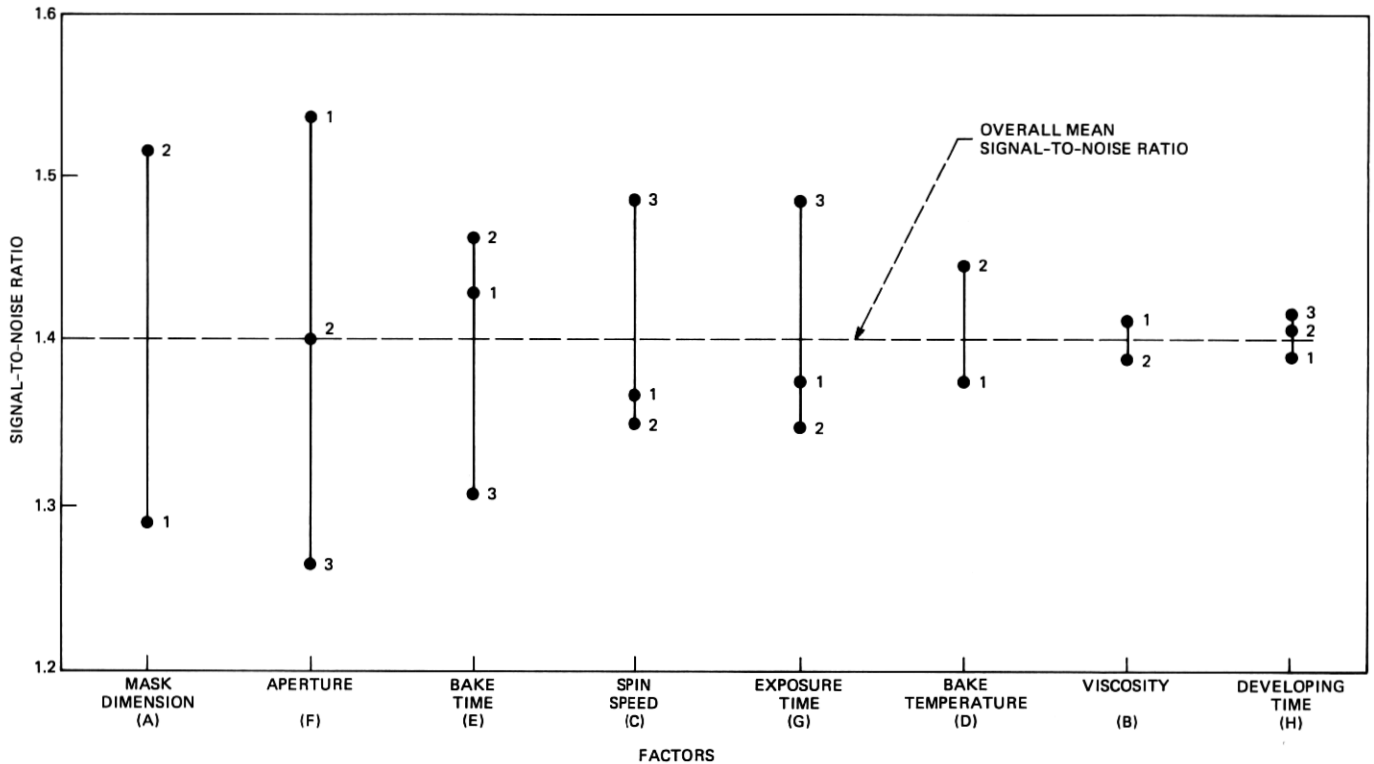

Fig. 3—Signal-to-noise ratios for pre-etch line width. The average s/n for each factor level is indicated by a dot. The number next to the dot indicates the factor level.

now x_i stands for the sum of the main effects of only A, E, and F. The F ratios, computed by dividing the factor mean square by the error mean square, are also shown in Table VIb. Factors A and F are significant using F-table values for the 5-percent significance level. So the mask dimension and the aperture are the control factors.

In performing the analysis of variance, we have tacitly assumed that the response for each experiment, here the s/n, has a normal distribution with constant variance. We are presently investigating the distributional properties of the s/n and their impact on the analysis of variance. In this paper we treat the significance levels as approximate.

The engineering significance of a statistically significant factor can be measured in terms of the percent contribution, a measure introduced by Taguchi.[3] The percent contribution is equal to the percent of the total sum of squares explained by that factor after an appropriate estimate of the error sum of squares has been removed from it. The larger the percent contribution, the more can be expected to be achieved by changing the level of that factor. Computation of the percent contribution is illustrated in Appendix B, and the results are shown in Table VIb.

From Table VIb we see that both the factors A (mask dimension) and F (aperture) contribute in excess of 20 percent each to the total

Table VI—Pre-etch line width

(a) ANOVA for s/n

	Source	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square
A	Mask Dimension	1	0.2399	0.2399	$\sigma^2 + \phi_A$
BD	Viscosity Bake Temperature	2	0.0169	0.0085	$\sigma^2 + \phi_{BD}$
C	Spin Speed	2	0.0668	0.0334	$\sigma^2 + \phi_C$
E	Bake Time	2	0.0804	0.0402	$\sigma^2 + \phi_E$
F	Aperture	2	0.2210	0.1105	$\sigma^2 + \phi_F$
G	Exposure Time	2	0.0634	0.0317	$\sigma^2 + \phi_G$
H	Developing Time	2	0.0017	0.0009	$\sigma^2 + \phi_H$
Error		4	0.1522	0.0381	σ^2
Total		17	0.8423		

(b) Pooled ANOVA for s/n

	Source	Degrees of Freedom	Sum of Squares	Mean Square	F	Percent Contribution
A	Mask Dimension	1	0.2399	0.2399	9.56*	25.5
E	Bake Time	2	0.0804	0.0402	1.60	3.6
F	Aperture	2	0.2210	0.1105	4.40*	20.3
Error		12	0.3010	0.0251		50.6
Total		17	0.8423			100.00

$F_{1,12}(0.95) = 4.75$.
$F_{2,12}(0.95) = 3.89$.
* Factors significant at 95-percent confidence level.

sum of squares. So the factors A and F are not only statistically significant, they have a sizable influence on the s/n. These results are consistent with Fig. 3. They will be used in Section IX for selecting the control factors.

7.3.2 Analysis of the means

Now we analyze the mean pre-etch line widths, $\bar{x}_i$ values, to find a signal factor.

The estimates of the mean line widths for all factor levels are given in Table VII. These estimates are graphically shown in Fig. 4. It is apparent that the levels of viscosity, mask dimension, and spin speed cause a relatively large change in the mean line width. Developing time and aperture have a small effect on the line width. The remaining two factors have an intermediate effect.

The linear model used to analyze this data is the same as eq. (1), except that now y_i stands for the mean pre-etch line width rather than the s/n.

The original and the pooled ANOVA tables for the mean pre-etch line width are given in Tables VIIIa and b, respectively. Because the design is not orthogonal with respect to the factors B and D, we need a special method, described in Appendix C, to separate S_{BD} into S_B and S_D.

It is clear from Table VIIIb that the mask dimension (A), viscosity (B), and spin speed (C) have a statistically significant effect on the mean pre-etch line width. Also, these factors together contribute more than 70 percent to the total sum of squares. These results will be used in Section IX for selecting the signal factor.

7.4 Post-etch line width

The analysis of the post-etch line-width data is similar to the analysis of the pre-etch line-width data. The mean, the standard deviation, and the s/n for each experiment are shown in Table IX.

Table VII—Pre-etch line width for the mean line width

		Mean Line Width (μm)		
	Factor	Level 1	Level 2	Level 3
A	Mask Dimension	2.39	2.87	
BD	Viscosity Bake Temperature	(B_1D_1) 2.83	(B_2D_1) 2.31	(B_1D_2) 2.74
B	Viscosity	2.79	2.31	
D	Bake Temperature	2.57	2.74	
C	Spin Speed	2.40	2.59	2.89
E	Bake Time	2.68	2.68	2.53
F	Aperture	2.68	2.56	2.64
G	Exposure Time	2.74	2.66	2.49
H	Developing Time	2.60	2.60	2.69

Overall mean line width = 2.63 μm.

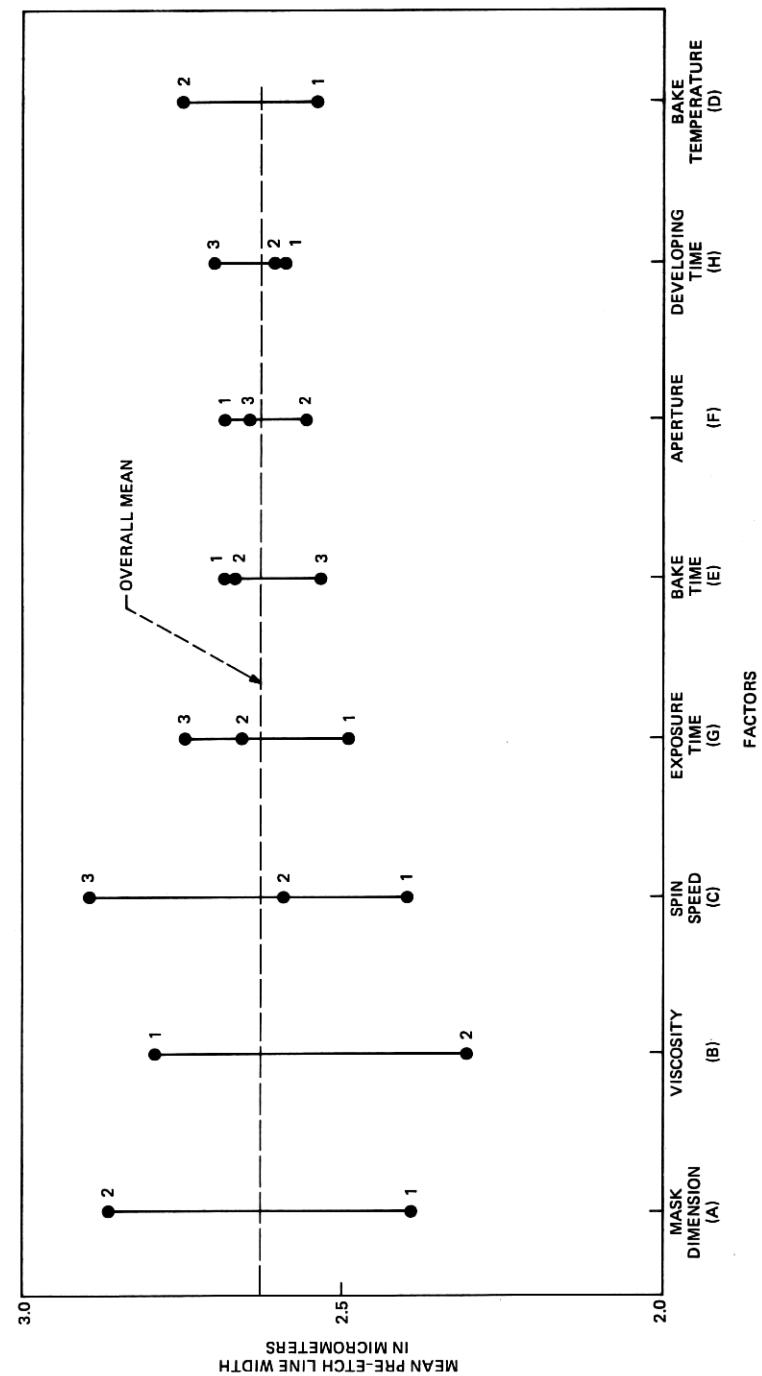

Fig. 4—Mean pre-etch line width. The mean line width for each factor level is indicated by a dot. The number next to each dot indicates the factor level.

Table VIII—Pre-etch line width

(a) ANOVA for mean line width

	Source	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square
A	Mask Dimension	1	1.05	1.050	$\sigma^2 + \phi_A$
BD	Viscosity Bake Temperature	2	0.95	0.475	$\sigma^2 + \phi_{BD}$
C	Spin Speed	2	0.73	0.365	$\sigma^2 + \phi_C$
E	Bake Time	2	0.10	0.050	$\sigma^2 + \phi_E$
F	Aperture	2	0.05	0.025	$\sigma^2 + \phi_F$
G	Exposure Time	2	0.19	0.095	$\sigma^2 + \phi_G$
H	Developing Time	2	0.04	0.020	$\sigma^2 + \phi_H$
Error		4	0.26	0.065	σ^2
Total		17	3.37		

(b) Pooled ANOVA for mean line width

	Source	Degrees of Freedom	Sum of Squares	Mean Square	F	Contribution Percent
A	Mask Dimension	1	1.05	1.050	19.81*	29.6
B	Viscosity	1	0.83	0.834	15.74*	22.6
C	Spin Speed	2	0.73	0.365	6.89*	18.5
G	Exposure Time	2	0.19	0.095	1.79	2.5
Error		11	0.58	0.053		26.8
Total		17	3.37			100.0

$F_{1,11}(0.95) = 4.84$.
$F_{2,11}(0.95) = 3.98$.
* Factors significant at 95-percent confidence level.

Table IX—Post-etch line-width data

Experiment Number	Mean Line Width, $\bar{x}$ (μm)	Standard Deviation of Line Width s (μm)	s/n $\eta = \log(\bar{x}/s)$
1	2.84	0.11	1.42
2	3.14	0.063	1.70
3	3.65	0.15	1.40
4	2.57	0.20	1.11
5	1.72	0.40	0.63
6	3.12	0.27	1.07
7	2.62	0.19	1.14
8	4.10	0.18	1.37
9	3.55	0.26	1.13
10	3.31	0.35	0.98
11	3.60	0.38	0.98
12	4.17	0.27	1.18
13	3.10	0.16	1.29
14	3.14	0.16	1.29
15	2.55	0.21	1.09
16	2.81	0.37	0.88
17	3.78	0.22	1.23
18	4.34	0.078	1.75

The average s/n and the mean line width for each factor level are shown in Tables Xa and b, respectively.

The linear model (1) was again used to analyze the post-etch line-width data. The ANOVA for the signal-to-noise ratios, Table XIa, indicates that none of the nine process factors has a significant effect (approximately 5-percent level) on the s/n for the post-etch line width. The pooled ANOVA for the mean post-etch line widths is shown in Table XIb. It is obvious from the table that the viscosity, exposure, spin speed, mask dimension, and developing time have significant effects (5-percent level) on the mean line width. The contribution of these factors to the total sum of squares exceeds 90 percent. The mean line width for each factor level is shown graphically in Fig. 5.

VIII. ANALYSIS OF POST-ETCH WINDOW-SIZE DATA

Some windows are printed and open while the others are not. Thus the window-size data are mixed categorical-continuous in nature. Analysis of such data is done by converting all the data to the categorical type and then using the 'accumulation analysis' method, which is

Table X—Post-etch line width

(a) Average signal-to-noise ratios

Factor	Average s/n		
	Level 1	Level 2	Level 3
A Mask Dimension	1.22	1.19	
BD Viscosity Bake Temperature	(B_1D_1) 1.28	(B_2D_1) 1.08	(B_1D_2) 1.25
B Viscosity	1.27	1.08	
D Bake Temperature	1.18	1.25	
C Spin Speed	1.14	1.20	1.27
E Bake Time	1.16	1.28	1.17
F Aperture	1.28	1.22	1.11
G Exposure Time	1.26	1.33	1.02
H Developing Time	1.09	1.20	1.32
I Etch Time	1.21	1.18	1.23

Overall average s/n = 1.205

(b) Mean line width

Factor	Mean Line Width (μm)		
	Level 1	Level 2	Level 3
A Mask Dimension	3.03	3.42	
BD Viscosity Bake Temperature	(B_1D_1) 3.45	(B_2D_1) 2.70	(B_1D_2) 3.53
B Viscosity	3.49	2.70	
D Bake Temperature	3.08	3.53	
C Spin Speed	2.88	3.25	3.56
E Bake Time	3.15	3.18	3.35
F Aperture	3.28	3.22	3.18
G Exposure Time	3.52	3.34	2.83
H Developing Time	3.04	3.09	3.56
I Etch Time	3.14	3.22	3.32

Overall mean line width = 3.23 μm.

described by Taguchi in Refs. 3 and 8. Factors that are found significant in this analysis are control factors.

The window sizes were divided into the following five categories:

Category	Description (micrometers)
I	Window not open or not printed
II	(0, 2.25)
III	[2.25, 2.75)
IV	[2.75, 3.25]
V	(3.25, ∞)

Note that these categories are ordered with respect to window size. The target window size at the end of step (*vi*) was 3 μm. Thus category IV is the most desired category, while category I is the least desired category. Table XII summarizes the data for each of the experiments by categories. To simplify our analysis, we shall presume that a missing wafer has the same readings as the observed wafer for that experiment. This is reflected in Table XII, where we show the combined readings for the two wafers of each experiment.

Table XI—Post-etch line width

a) ANOVA for s/n

	Source	Degrees of Freedom	Sum of Squares	Mean Square	F
A	Mask Dimension	1	0.005	0.005	0.02
B	Viscosity	1	0.134	0.134	0.60
D	Bake Temperature	1	0.003	0.003	0.01
C	Spin Speed	2	0.053	0.027	0.12
E	Bake Time	2	0.057	0.028	0.13
F	Aperture	2	0.085	0.043	0.19
G	Exposure Time	2	0.312	0.156	0.70
H	Developing Time	2	0.156	0.078	0.35
I	Etch Time	2	0.008	0.004	0.02
Error		2	0.444	0.222	
Total		17	1.257		

b) Pooled ANOVA for mean line width

	Source	Degrees of Freedom	Sum of Squares	Mean Square	F	Percent Contribution
A	Mask Dimension	1	0.677	0.677	16.92*	8.5
B	Viscosity	1	2.512	2.512	63.51*	32.9
C	Spin Speed	2	1.424	0.712	17.80*	17.9
G	Exposure Time	2	1.558	0.779	19.48*	19.6
H	Developing Time	2	0.997	0.499	12.48*	12.2
Error		9	0.356	0.040		8.9
Total		17	7.524			100.0

$F_{1,9}(0.95) = 5.12$.
$F_{2,9}(0.95) = 4.26$.
* Factors significant at 95-percent confidence level.

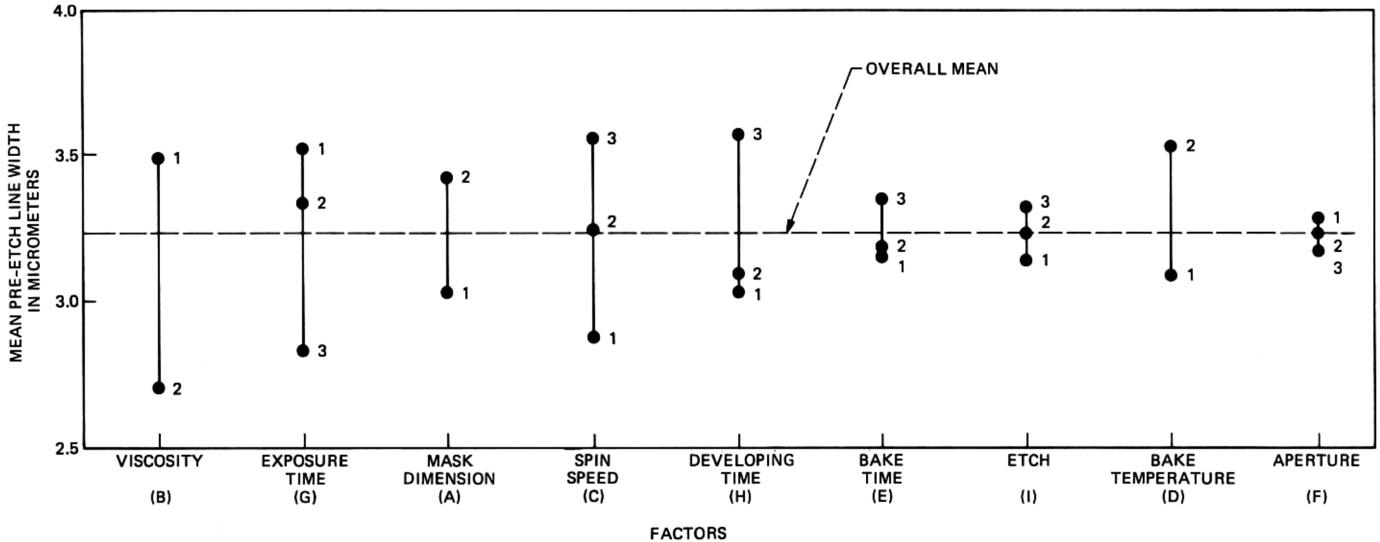

Fig. 5—Mean post-etch line width. The mean line width for each factor level is indicated by a dot. The number next to a dot indicates the factor level.

Table XII—Post-etch window-size data—frequencies by experiment

Experiment No.	Frequency Distribution for Wafer 1					Frequency Distribution for Wafer 2					Combined Frequency for the Two Wafers				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
1	5	0	0	0	0	5	0	0	0	0	10	0	0	0	0
2	0	1	0	2	2	0	2	3	0	0	0	3	3	2	2
3	0	0	0	4	0	1	0	0	3	0	1	0	0	9	0
4	5	0	0	0	0	5	0	0	0	0	10	0	0	0	0
5	*	*	*	*	*	5	0	0	0	0	10	0	0	0	0
6	0	3	2	0	0	5	0	0	0	0	5	3	2	0	0
7	5	0	0	0	0	5	0	0	0	0	10	0	0	0	0
8	5	0	0	0	0	0	0	0	5	0	5	0	0	5	0
9	0	0	0	5	0	0	1	4	0	0	0	1	4	5	0
10	0	5	0	0	0	2	0	3	0	0	2	5	3	0	0
11	0	0	0	5	0	1	1	2	1	0	1	1	2	6	0
12	1	0	1	3	0	0	0	0	0	4	1	0	1	3	5
13	0	0	3	2	0	5	0	0	0	0	5	0	3	2	0
14	1	3	1	0	0	5	0	0	0	0	6	3	1	0	0
15	5	0	0	0	0	*	*	*	*	*	10	0	0	0	0
16	5	0	0	0	0	5	0	0	0	0	10	0	0	0	0
17	0	0	0	3	2	0	0	4	0	1	0	0	4	3	3
18	0	0	0	0	5	*	*	*	*	*	0	0	0	0	10

* Implies data missing.

Table XIII gives the frequency distribution corresponding to each level of each factor. To obtain the frequency distribution for a specific level of a specific factor, we summed the frequencies of all the experiments that were conducted with that particular level of that particular factor. For example, the frequency distribution for the first level of factor C (low spin speed) was obtained by summing the frequency distributions of experiments with serial numbers 1, 4, 7, 10, 13, and 16. These six experiments were conducted with level 1 of factor C.

The frequency distributions of Table XIII are graphically displayed by star plots in Fig. 6. From this figure and the table it is apparent that a change in the level of viscosity, spin speed, or mask dimension causes a noticeable change in the frequency distribution. A change in the level of etch time, bake time, or bake temperature seems to have only a small effect on the frequency distribution. The effects of the other factors are intermediate.

We now determine which factors have a significant effect on the frequency distribution of the window sizes. The standard chi-square test for multinomial distributions is not appropriate here because the categories are ordered. The accumulation analysis method has an intuitive appeal and has been empirically found by Professor Taguchi to be effective in analyzing ordered categorical data. The method consists of the following three steps:

(*i*) Compute the cumulative frequencies. Table XIII shows the cumulative frequencies for all factor levels. The cumulative categories are denoted with parentheses. Thus (III) means sum of categories I,

Table XIII—Post-etch window-size data—frequencies by factor level

Factor Levels	Frequencies					Cumulative Frequencies				
	I	II	III	IV	V	(I)	(II)	(III)	(IV)	(V)
Mask Dimension										
A_1	51	7	9	21	2	51	58	67	88	90
A_2	35	9	14	14	18	35	44	58	72	90
Viscosity, Bake Temperature										
B_1D_1	15	9	9	20	7	15	24	33	53	60
B_2D_1	46	6	6	2	0	46	52	58	60	60
B_1D_2	25	1	8	13	13	25	26	34	47	60
Spin Speed										
C_1	47	5	6	2	0	47	52	58	60	60
C_2	22	7	10	16	5	22	29	39	55	60
C_3	17	4	7	17	15	17	21	28	45	60
Bake Time										
E_1	31	2	10	14	3	31	33	43	57	60
E_2	26	3	7	7	17	26	29	36	43	60
E_3	29	11	6	14	0	29	40	46	60	60
Aperture										
F_1	32	7	5	6	10	32	39	44	50	60
F_2	36	3	4	10	7	36	39	43	53	60
F_3	18	6	14	19	3	18	24	38	57	60
Exposure Time										
G_1	26	3	10	13	8	26	29	39	52	60
G_2	18	12	11	7	12	18	30	41	48	60
G_3	42	1	2	15	0	42	43	45	60	60
Developing Time										
H_1	37	4	6	8	5	37	41	47	55	60
H_2	27	11	12	5	5	27	38	50	55	60
H_3	22	1	5	22	10	22	23	28	50	60
Etch Time										
I_1	37	5	3	5	10	37	42	45	50	60
I_2	21	8	14	15	2	21	29	43	58	60
I_3	28	3	6	15	8	28	31	37	52	60
Totals	86	16	23	35	20	86	102	125	160	180

II, and III. Note that the cumulative category (V) is the same as the total number of window readings for the particular factor level.

(*ii*) Perform "binary data" ANOVA[7] on each cumulative category except the last category, viz. (V). Note that a certain approximation is involved in the significance levels suggested by this ANOVA because the observations are not normally distributed.

(*iii*) Assign weights to each cumulative category. These weights are inversely proportional to the Bernoulli trial variance. Let cum_c be the total number of windows in the cumulative category, c, as given in the bottom row of Table XIII. Then the weight for that category is:

$$W_c = \frac{1}{\frac{cum_c}{180} \times \left(1 - \frac{cum_c}{180}\right)} = \frac{180^2}{cum_c(180 - cum_c)}.$$

These weights are shown in Appendix D for each category.

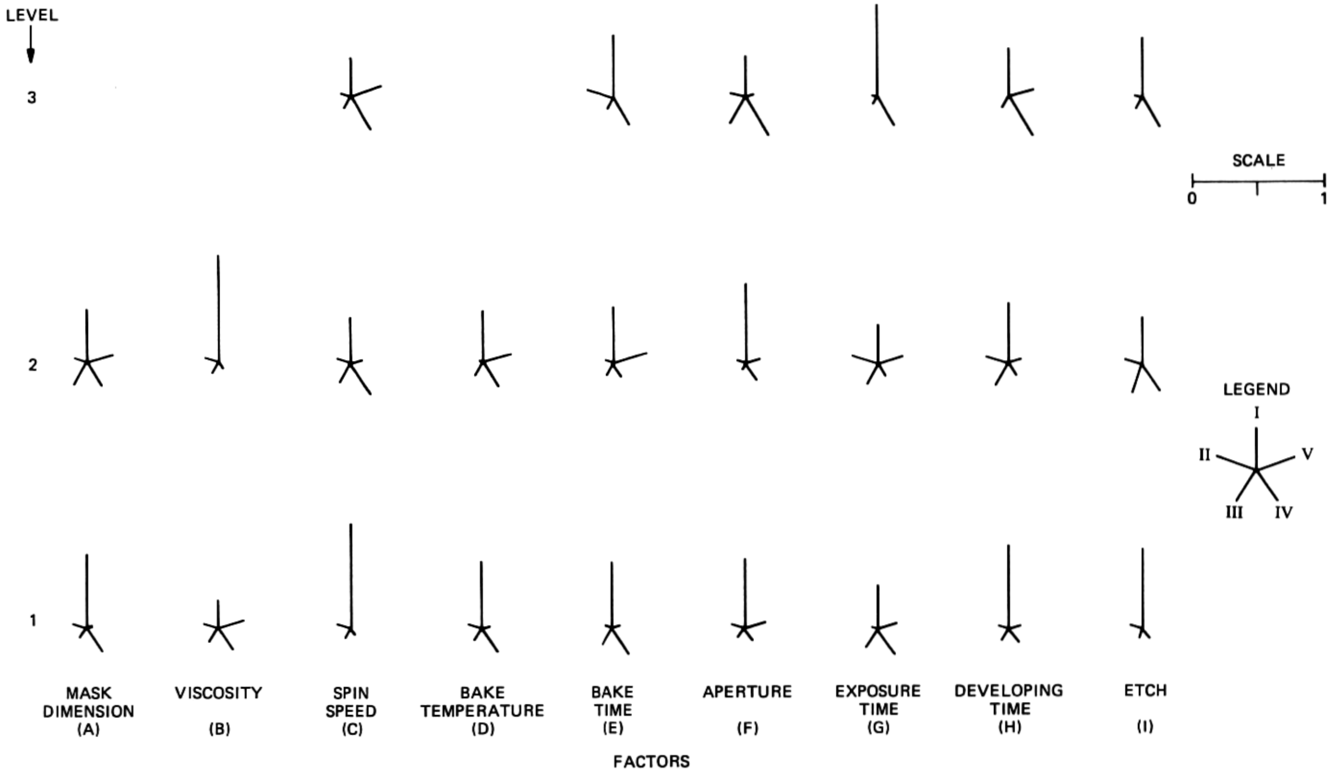

Fig. 6—Post-etch window size. The star gives the frequency distribution for each factor level.

Then for each factor and for each error term the accumulated sum of squares is taken to be equal to the weighted sum of the sum of squares for all cumulative categories.

The intuitive appeal for accumulation analysis is that by taking cumulative frequencies we preserve the order of the categories. By giving weights inversely proportional to the sampling errors in each cumulative category, we make the procedure more sensitive to a change in the variance. The difficulty is that the frequencies of the cumulative categories are correlated. So the true level of significance of the F-test may be somewhat different from that indicated by the F table. More work is needed to understand the statistical properties of the accumulation analysis.

Table XIV gives the final ANOVA with accumulated sum of squares. The computations are illustrated in Appendix D. For each cumulative category, the following nested, mixed linear model was used in performing the ANOVA:

$$y_{ijk} = \mu + x_i + e_{1ij} + e_{2ijk}, \tag{2}$$

where

i $= 1, \cdots, 18$ stands for the experiment.

j $= 1, 2$ stands for wafer within the experiment.

k $= 1, \cdots, 5$ stands for replicate or position within wafer within the experiment.

μ is the overall mean.

x_i is the fixed effect of the factor-level combination used in experiment i. Here we consider only the main effect for each of the factors. See the discussion of model (1) in Section 7.3 for more details of the interpretation of x_i.

e_{1ij} is the random effect for wafer j within experiment i.

e_{2ijk} is the random error for replicate k within wafer j within experiment i.

y_{ijk} is the observation for replicate k in wafer j in experiment i. y_{ijk} takes a value 1 if the window size belongs to the particular category. Otherwise, the value is zero.

The expected mean squares for this ANOVA model are also shown in Table XIV. The variances of e_1 and e_2 are denoted by σ_1^2 and σ_2^2, respectively. The effects of the factors A through I are denoted by ϕ with an appropriate subscript. The effect of lack of fit is denoted by ϕ_ℓ. We assume that the random variables e_{1ij} and e_{2ijk} are independent for all values of i, j and k. The degrees of freedom shown in Table XIV have been adjusted for the fact that three experiments have only one wafer each.

For testing the significance of the effect of error between wafers within experiments, the relevant denominator sum of squares is the

Table XIV—Post-etch window size

ANOVA for accumulation analysis

Source	Degrees of Freedom	Sum of Squares	Mean Square	F	(Expected Mean Square) ÷ $\bar{W}$
A Mask Dimension	4	26.64	6.66	2.67*	$\sigma_2^2 + 5\sigma_1^2 + \phi_A$
BD Viscosity-Bake Temperature	8	112.31	14.04	5.64*	$\sigma_2^2 + 5\sigma_1^2 + \phi_{BD}$
C Spin Speed	8	125.52	15.69	6.30*	$\sigma_2^2 + 5\sigma_1^2 + \phi_C$
E Bake Time	8	36.96	4.62	1.86	$\sigma_2^2 + 5\sigma_1^2 + \phi_E$
F Aperture	8	27.88	3.49	1.40	$\sigma_2^2 + 5\sigma_1^2 + \phi_F$
G Exposure Time	8	42.28	5.29	2.12*	$\sigma_2^2 + 5\sigma_1^2 + \phi_G$
H Developing Time	8	45.57	5.70	2.29*	$\sigma_2^2 + 5\sigma_1^2 + \phi_H$
I Etch Time	8	23.80	2.98	1.20	$\sigma_2^2 + 5\sigma_1^2 + \phi_I$
Lack of Fit	8	17.25	2.16	0.87	$\sigma_2^2 + 5\sigma_1^2 + \phi_l$
Error Between Wafers Within Experiment	60	149.33	2.49	11.69*	$\sigma_2^2 + 5\sigma_1^2$
Error Between Replicates Within Wafers Within Experiment	528	112.45	0.21		σ_2^2
Total	656	720.00			

$\bar{W} = (W_{(I)} + W_{(II)} + W_{(III)} + W_{(IV)})/4.$
$F_{4,60}(0.95) = 2.53$, $F_{8,60}(0.95) = 2.10$, $F_{60,528}(0.95) = 1.32$.

(b) Separation of S_{BD}

Source		Degrees of Freedom	Sum of Squares	Mean Square	F
B	Viscosity	4	87.38	21.85	8.78*
D	Bake Temperature	4	6.55	1.64	0.66

* Factors significant at 95-percent confidence level.

estimate of σ_2^2. The corresponding F value is 11.69, which is significant far beyond the nominal 5-percent level. To test for the lack of fit of the main-effects-only model, the appropriate denominator is the estimate $\sigma_2^2 + 5\sigma_1^2$. The corresponding F ratio is 0.87. This indicates that the main-effects-only model adequately describes the observed data relative to the random errors between wafers. For testing the significance of the process factors, the denominator mean square is again the estimate of $\sigma_2^2 + 5\sigma_1^2$. We see that the mask dimension, viscosity, spin speed, exposure time, and developing time have a significant effect (approximately 5-percent level) on the window size. The effects of the other factors are not significant.

IX. SELECTION OF OPTIMUM FACTOR LEVELS

The following table summarizes the significant results of the analyses performed in Sections VII and VIII. In each category, the factors are arranged in descending order according to the F value.

Significant effect on s/n:

Pre-etch line width: A, F
Post-etch line width: None
Significant effect on mean:
Pre-etch line width: A, B, C
Post-etch line width: B, G, C, A, H
Significant factors identified by accumulation analysis:
Post-etch window size: B, C, A, H, G

Factors that have a significant effect on the s/n and the factors identified to be significant by the accumulation analysis are all control factors. Setting their levels equal to optimum levels minimizes the process variability. Here the control factors are A, F, B, C, H, and G.

To keep the process mean on target we use a signal factor. Ideally, the signal factor should have a significant effect on the mean, but should have no effect on the s/n. Then changing the level of the signal factor would affect only the mean. In practice, a small effect on the s/n may have to be tolerated.

Among the factors (A, B, C, G, and H) that have a significant effect on the mean, factors A, B, and C are relatively strong control factors as measured by the F statistics for the accumulation analysis and the ANOVA for pre-etch line-width s/n. Also, these factors are relatively difficult to change during production. So A, B, and C are not suitable as signal factor. Between the remaining two factors, G and H, G has greater effect on the mean and also shows as a less significant factor in accumulation analysis. So exposure time was assigned to be the signal factor.

The optimum levels for the control factors were selected as follows. The mask dimension (A) and the aperture (F) have a significant effect on the s/n for pre-etch line width. From Table V we see that the 2.5-μm mask (level 2) has a higher s/n than the 2.0-μm mask. Hence 2.5 μm was chosen to be the optimum mask dimension. Also, aperture 1 (level 1) has the highest s/n among the three apertures studied. However, because of the past experience, aperture 2 was chosen to be the preferred level.

The accumulation analysis of the post-etch window-size data indicated that the viscosity, spin speed, mask dimension, developing time, and exposure have statistically significant effects on the frequency distribution. The optimum levels of these factors can be determined from Table XIII and Fig. 6 to be those that have the smallest fraction of windows not open (category I) and the largest fraction of windows in the range 3.0 ± 0.25 μm (category IV). Because it is more critical to have all the windows open, when there was a conflict we took the requirement on category I to be the dominant requirement. The optimum levels are: 2.5-μm mask dimension, viscosity 204, 4000-rpm spin speed, 60-second developing time, and normal exposure.

Table XV—Optimum factor levels

Label	Factors Name	Standard Levels	Optimum Levels
A	Mask Dimension (μm)	2.0	2.5
B	Viscosity	204	204
C	Spin Speed (rpm)	3000	4000
D	Bake Temperature (°C)	105	105
E	Bake Time (min)	30	30
F	Aperture	2	2
G	Exposure (PEP setting)	Normal	Normal
H	Developing Time (s)	45	60
I	Plasma Etch Time (min)	13.2	13.2

Table XV shows side by side the optimum factor levels and the standard levels as of September 1980. Note that our experiment has indicated that the mask dimension be changed from 2.0 μm to 2.5 μm, spin speed from 3000 rpm to 4000 rpm, and developing time from 45 seconds to 60 seconds. The exposure time is to be adjusted to get the correct mean value of the line width and the window size. The levels of the other factors, which remain unchanged, have been confirmed to be optimum to start with.

In deriving the optimum conditions we have conducted a highly fractionated factorial experiment and have considered only the main effects of the factors. The interactions between the factors have been ignored. If the interactions are strong compared to the main effects, then there is a possibility that the optimum conditions thus derived would not improve the process. So experiments have to be conducted to verify the optimum conditions. The verification was done in conjunction with the implementation, which is described next.

X. IMPLEMENTATION AND THE BENEFITS OF THE OPTIMUM LEVELS

We started to use the optimum process conditions given in Table XV in the Integrated Circuits Design Capability Laboratory in January 1981. In the beginning the exposure was set at 90, which is the normal setting given in Table I. We observed that the final window at the end of step (*xi*) was much larger than the target size of 3.5 μm. Through successive experiments, we reduced the exposure time until the mean final window size came to about 3.5 μm. The corresponding exposure setting is 140. Since then the process has been run at these conditions. The benefits of running the process at these conditions are:

(*i*) The pre-etch line width is routinely used as a process quality indicator. Before September 1980 the standard deviation of this indicator was 0.29 μm on a base line chip (DSO chip). With the optimum process parameters, that standard deviation has come down to 0.14 μm. This is a two-fold reduction in standard deviation, or a four-fold reduction in variance. This is evidenced by Fig. 7, which shows a typical photograph of the PLA area of a *BELLMAC*-32 microprocessor chip fabricated by the new process. Note that windows in Fig. 7 are much more uniform in size compared to the windows in Fig. 2. Also, all windows are printed and open in Fig. 7.

(*ii*) After the final step of window forming, i.e., after step (*xi*), the windows are visually examined on a routine basis. Analysis of the quality control data on the DSO chip, which has an area of approximately 0.19 cm^2, showed that prior to September 1980 about 0.12 window per chip was either not open or not printed (i.e., approximately one incidence of window not open or not printed was found in eight chips). With the new process only 0.04 window per chip is not open or printed (i.e., approximately one incidence of window not open or printed is found in twenty-five chips). This is a three-fold reduction in defect density due to unopened windows.

(*iii*) Observing these improvements over several weeks, the process engineers gained a confidence in the stability and robustness of the new process parameters. So they eliminated a number of in-process checks. As a result the overall time spent by the wafers in window photolithography has been reduced by a factor of two.

The optimum parameter levels were first used in the Integrated Circuit Device Capability Laboratory with only a few codes of ICs. Subsequently, these parameter levels were used with all codes of 3.5-μm technology chips, including *BELLMAC*-4 microcomputer and *BELLMAC*-32 microprocessor chips. The mask dimension change from 2.0 to 2.5 μm is now a standard for 3.5-μm CMOS technology.

XI. DISCUSSION AND FUTURE WORK

The off-line quality control method is an efficient method of improving the quality and the yield of a production process. The method has a great deal of similarity with the response surface method[9] and the evolutionary operations method,[10] which are commonly known in statistical literature in this country. Both the response surface and the evolutionary operations methods are used to maximize the yield of a production process and they both make use of the experimental design techniques. The main difference is that in the off-line quality control method the process variability that has a great impact on the product quality is the objective function. In the response surface and evolutionary operations methods, the process variability is generally not considered. Thus, intuitively, the optimum levels derived by using the off-line quality control method can be expected to be more robust, stable, and dependable.

In the response surface method one typically uses a relatively large fraction of the factorial experiment. However, in off-line quality control usually a very small fraction is chosen. Another difference is that in

Fig. 7—Typical results for contact windows in PLA area (January through March, 1981) using parameters derived from this experiment. The contact windows are round shaped.

the response surface method the objective function is considered to be a continuous function approximated by a low-order polynomial. In off-line quality control, we can simultaneously study both the continuous and discrete factors.

Our application of the off-line quality control method to the window-cutting process in the Murray Hill 3.5-μm CMOS technology, as seen from the earlier sections, has resulted in improved control of window size, lower incidence of unopened windows, and reduced time for window photolithography. Presently, we have undertaken to optimize two more steps in IC fabrication. Those steps are polysilicon patterning and aluminum patterning. Both these processes, like the window-cutting process, involve photolithography and are among the more critical processes of IC fabrication. We think that the method has a great potential and would like to see applications in various parts of Bell Laboratories and Western Electric Company.

XII. ACKNOWLEDGMENTS

The authors acknowledge with gratitude the support and the encouragement given by R. L. Chaddha, R. Edwards, J. V. Dalton, A. B. Hoadley, and G. R. Weber. R. G. Brandes, K. J. Orlowski, and members of the IC Process Operations Group provided valuable help in conducting the experiments. We also thank Ramon Leon for helpful discussions on the graphical representation of the data and I. J. Terpenning for helpful comments on an earlier draft of this paper. The

original manuscript was carefully prepared by Jill Cooper on the *UNIX** operating system.

REFERENCES

1. G. Taguchi, "Off-line and On-line Quality Control Systems," International Conference on Quality Control, Tokyo, Japan, 1978, pp. B4-1–5.
2. R. N. Kackar and M. S. Phadke, unpublished work.
3. G. Taguchi and Y. Wu, *Introduction to Off-Line Quality Control*, Central Japan Quality Control Association, Nagoya, Japan, 1980.
4. A. B. Glaser and G. A. Subak-Sharp, *Integrated Circuits Engineering*, Reading, MA: Addison-Wesley, 1977.
5. W. Q. Meeker, G. H. Hahn, and P. I. Feder, "A Computer Program for Evaluating and Comparing Experimental Designs and Some Applications," American Statistician, *29*, No. 1 (February 1975), pp. 60–4.
6. C. R. Hicks, *Fundamental Concepts in the Design of Experiments*, New York, NY: Holt, Rinehart and Winston, 1973.
7. S. R. Searle, *Linear Models*, New York, NY: John Wiley, 1971.
8. G. Taguchi, "A New Statistical Analysis Method for Clinical Data, the Accumulation Analysis, in Contrast with the Chi-square Test," Shinjuku Shobo, Tokyo, Japan, 1975.
9. R. H. Myers, *Response Surface Methodology*, Newton, MA: Allyn and Bacon, 1971.
10. G. E. P. Box and N. R. Draper, *Evolutionary Operations: A Statistical Method for Process Improvement*, New York, NY: John Wiley, 1969.
11. G. W. Snedecor and W. G. Cochran, *Statistical Methods*, Ames, Iowa: Iowa State University Press, 1980.

APPENDIX A

Computation of the Sum of Squares—Analysis of the Signal-to-Noise Ratio for Pre-Etch

The computations of the sum of squares tabulated in Table VIa are illustrated below.

$$S_m = \text{Correction Factor}$$

$$= \frac{\left(\sum_{i=1}^{18} \eta_i\right)^2}{18} = \frac{(25.2202)^2}{18} = 35.3366$$

$$S_A = \text{Sum of squares for factor A}$$

$$= \frac{(9m_{A_1})^2 + (9m_{A_2})^2}{9} - S_m$$

$$= \frac{(11.5711)^2 + (13.6491)^2}{9} - 35.3366$$

$$= 0.2399 \qquad (d.f. = 1)$$

$$S_C = \frac{(6m_{C_1})^2 + (6m_{C_2})^2 + (6m_{C_3})^2}{6} - S_m$$

* Trademark of Bell Laboratories.

$$= \frac{(8.1979)^2 + (8.1017)^2 + (8.9206)^2}{6} - S_m$$

$$= 0.0668 \qquad (d.f. = 2).$$

Sums of squares for the factors E, F, G, and H were calculated similarly. The combined sum of squares due to B and D is given by

$$S_{BD} = \text{Sum of squares for the column BD}$$

$$= \frac{(6m_{B_1D_1})^2 + (6m_{B_1D_2})^2 + (6m_{B_2D_1})^2}{6} - S_m$$

$$= \frac{(8.2524)^2 + (8.6649)^2 + (8.3029)^2}{6} - 35.3366$$

$$= 0.0169 \qquad (d.f. = 2).$$

The total sum of squares is

$$S_T = \sum_{i=1}^{18} \eta_i^2 - S_m = 0.8423. \qquad (d.f. = 17).$$

The error sum of squares is calculated by subtraction.

$$S_e = S_T - (S_A + S_{BD} + S_C + S_E + S_F + S_G + S_H)$$

$$= 0.1522 \qquad (d.f. = 4).$$

Here we do not compute the sum of squares due to factor I (etch time), because it has no influence on the pre-etch line width.

APPENDIX B

Computation of the Percent Contribution—Analysis of the Signal-to-Noise Ratio for Pre-Etch Line Width

The computation of the percent contribution is explained below. The contribution of factor A to the total sum of squares

$$= S_A - (d.f. \text{ of A})(\text{error mean square}).$$

Hence, the percent contribution for factor A

$$= \frac{S_A - (\text{d.f. of A})(\text{error mean square})}{\text{total sum of squares}} \times 100$$

$$= \frac{0.2399 - 0.0251}{0.8423} \times 100 = 25.5\%.$$

The percent contributions of E and F are determined similarly. Now consider, the contribution of error to the total sum of squares:

$$= S_e + (\text{total } d.f. \text{ for factors})(\text{error mean square}).$$

Hence, the percent contribution for error

$$= \frac{S_e + (\text{total } d.f. \text{ for factors})(\text{error mean square})}{\text{total sum of squares}} \times 100$$

$$= \frac{0.3010 + 5 \times 0.0251}{0.8423} \times 100 = 50.6\%.$$

APPENDIX C

Separation of S_{BD} into S_B and S_D—Analysis of the Mean Pre-Etch Line Width

The sum of squares, S_{BD}, can be decomposed in the following two ways[11] to obtain the contributions of the factors B and D:

$$S_{BD} = S'_{B(D)} + S'_D$$

and

$$S_{BD} = S'_{D(B)} + S'_B.$$

Here $S'_{B(D)}$ is the sum of squares due to B, assuming D has no effect; S'_D is the sum of squares due to D after eliminating the effect of B. The terms $S'_{D(B)}$ and S'_B are interpreted similarly. We have

$$S'_{B(D)} = \frac{(6m_{B_1D_1} + 6m_{B_1D_2} - 12m_{B_2D_1})^2}{(1^2 + 1^2 + 2^2) \times 6} = 0.903 \qquad (d.f. = 1)$$

$$S'_D = S_{BD} - S'_{B(D)} = 0.047 \qquad (d.f. = 1).$$

Similarly,

$$S'_{D(B)} = \frac{(6m_{B_1D_1} + 6m_{B_2D_1} - 12m_{B_1D_2})^2}{(1^1 + 1^2 + 2^2) \times 6} = 0.116 \qquad (d.f. = 1)$$

$$S'_B = S_{BD} - S'_{D(B)} = 0.834 \qquad (d.f. = 1).$$

For testing the significance of the factors B and D we use S'_B and S'_D, respectively. Note that S'_B and S'_D do not add up to S_{BD}, which is to be expected because the design is not orthogonal with respect to the factors B and D.

APPENDIX D

Computation of the Sum of Squares for Accumulation Analysis—Analysis of Post-Etch Window Size

The weights for the cumulative categories (I), (II), (III), and (IV) are given below. The frequencies of the bottom line of Table XIII are used in computing these weights. Therefore:

$$W_{(\mathrm{I})} = \frac{1}{\dfrac{86}{180} \times \dfrac{180 - 86}{180}} = 4.008$$

$$W_{(II)} = \frac{180^2}{102 \times (180 - 102)} = 4.072$$

$$W_{(III)} = \frac{180^2}{125 \times (180 - 125)} = 4.713$$

$$W_{(IV)} = \frac{180^2}{160 \times (180 - 160)} = 10.125.$$

Computation of the sum of squares tabulated in Table XIV are illustrated below:

$$S_A = W_{(I)} \times \left(\frac{51^2 + 35^2}{90} - \frac{86^2}{180}\right) + W_{(II)} \times \left(\frac{58^2 + 44^2}{90} - \frac{102^2}{180}\right)$$

$$+ W_{(III)} \times \left(\frac{67^2 + 58^2}{90} - \frac{125^2}{180}\right) + W_{(IV)} \times \left(\frac{88^2 + 72^2}{90} - \frac{160^2}{180}\right)$$

$$= 26.64,$$

and

$$S_C = W_{(I)} \times \left(\frac{47^2 + 22^2 + 17^2}{60} - \frac{86^2}{180}\right)$$

$$+ W_{(II)} \times \left(\frac{52^2 + 29^2 + 21^2}{60} - \frac{102^2}{180}\right)$$

$$+ W_{(III)} \times \left(\frac{58^2 + 39^2 + 28^2}{60} - \frac{125^2}{180}\right)$$

$$+ W_{(IV)} \times \left(\frac{60^2 + 55^2 + 45^2}{60} - \frac{160^2}{180}\right)$$

$$= 125.52.$$