Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

THE APPLICATION OF EXPERIMENTAL DESIGN TO THE ANALYSIS OF SEMICONDUCTOR MANUFACTURING LINES

Sarah J. Hood
Peter D. Welch

IBM Thomas J. Watson Research Center
Yorktown Heights, New York 10598

ABSTRACT

Flexible, interactive software is used to apply a two level fractional factorial design with ten factors to the discrete event simulation of a semiconductor manufacturing line. The sensitivity of cycle time to the number of tools and the operator to tool ratio at five tool groups that tend to have large queues preceding them is analyzed. The analysis confirms the proposed model for cycle time and illustrates the efficiency of using design of experiments for discrete event simulation.

1. INTRODUCTION

This paper presents the application of the design of experiments to the discrete event simulation of a complex manufacturing line. The discussion will use, as an example, a large model of an operating semiconductor line. The model contains on the order of one hundred tool groups processing multiple products. The flow is highly re-entrant, that is, jobs feed back through sequences of the tool groups up to fourteen times. The model includes tool breakdown and repair, preventative maintenance, rework, test wafer sendahead, and detailed operator schedules. The primary purpose of the model is to study control rules proposed for the line [Hood et al. 1989].

For the example discussed here, the line is assumed to be under constant load and in a stationary state. That is, the rate of output of product is assumed equal to the rate at which orders are inputted. No portion of the system is saturated. We are interested in the cycle time, the time from beginning of manufacture to completion, of one of the products. In particular we are interested in the sensitivity of this cycle time to the number of tools and the operator to tool ratio at five tool groups that tend to have more jobs queued than other tool groups. Thus there are ten factors in all.

Two level, fractional factorial designs were selected because of their efficiency and simplicity, and because of the broad range of design and analysis tools available for them. Box, Hunter and Hunter [1978] is recommended as a general reference. (In fact, it is interesting to note, that this reference includes an example of the application of two level designs to simulation. See Section 13.4.) The tool effects are labeled A, B, C, D and E and the corresponding operator effects V, W, X, Y and Z. In both cases a + effect corresponds to the larger number of tools or an operator to tool ratio of 1:1 and a − effect corresponds to the smaller number of tools or an operator to tool ratio of 1:2. Changing any of these effects will change the cycle time (waiting plus service time) at the corresponding tool group but would be expected to only negligibly interact with effects at other work stations. This follows from the fact that the system is, for the entire experiment, under the same load and is in stable, stationary states. Hence the output rates from each tool group do not change from experimental point to experimental point. The timing of the tool group outputs would change but the overall output rates would remain constant. This means that the input rates would also remain constant. Hence the only two level interactions one would expect to be significant would be AV, BW, CX, DY and EZ. Thus, since the overall cycle time is a linear combination of the tool group cycle times, we would expect a model of the form

$$R(\mu, x_1, \dots, x_5, y_1, \dots, y_5) = \mu + \sum_{i=1}^{5} \alpha_i x_i + \sum_{i=1}^{5} \beta_i y_i + \sum_{i=1}^{5} \gamma_i x_i y_i$$

where R is the mean cycle time for a particular run, μ is the mean cycle time over all runs, $x_i = \pm 1$, and $y_i = \pm 1$. The variables x_i and y_i reflect the levels, respectively, of the number of tools and the operator to tool ratio at the $i = 1, \dots, 5$ workstations. The coefficient γ_i represents the interaction between tools and operators at the *ith* workstation. The coefficients $\alpha_1, \dots, \alpha_5$, $\beta_1, \dots, \beta_5$, $\gamma_1, \dots, \gamma_5$ are equal to one half of the estimated effects A, ... , E, V, ... , Z, AV, ... , EZ respectively since they represent a unit change and the variables x_i and y_i take on the values ± 1. Hence the proposed model has 16 possible parameters. As noted earlier, in all cases $x_i = -1$ corresponds to the smaller number of tools and $x_i = +1$ corresponds to the larger number of tools; $y_i = -1$ corresponds to an operator to tool ratio of 1:2 and $y_i = +1$ corresponds to an operator to tool ratio of 1:1. Hence, *a priori*, we would expect $\alpha_1, \dots, \alpha_5$, $\beta_1, \dots, \beta_5$ all to be nonpositive, producing a reduction in cycle time. We had no expectations for the signs of the interactions, $\gamma_1, \dots, \gamma_5$.

2. THE DESIGN

To decide on an experimental design and later to analyze this design, an internal IBM graphical-statistical package, GRAFSTAT [Lane and Welch 1987] was used. It has a design of experiments component which allows the flexible design and analysis of two level fractional factorial experiments. The initial design was a resolution 3 design in which the interactions AV, BW, CX, DY and EZ were estimable. That is, a design where none of the main effects and the interactions AV, BW, CX, DY and EZ were confounded with one another; a design where, given that the model is true, unbiased estimates can be obtained for all the parameters. (A resolution 3 design is one in which none of the main effects are confounded with one another.) Such a design was obtained and required 16 runs. Thus, it was a 1/64 replicate design: a design requiring only 1/64 of the number of runs required for a complete study of the all 1024 possible combinations of the 10 factors. With 16 runs all the parameters of the model could be estimated but there were no degrees of freedom for estimating error because the state space spanned by the experiment and the state space spanned by the model were identical. Hence, confidence intervals could not be put on the parameters of the model and there was no capability for testing the assumptions of the model; namely that the five interactions AV, ... , EZ were the only significant ones.

Hence the design was expanded to a resolution 3 design with the interactions AV, BW, CX, DY and EZ estimable and with 32 runs. This design left 16 degrees of freedom to check for the accuracy of the model. The design matrix is given in Table 1. Of the 16 degrees of freedom associated with error: 14 are associated with two factor interactions, i.e. two factor interactions are in their alias sets; and 2 are associated with three factor interactions or higher. The alias structure (out to 2 factor interactions) for the 15 effects in the

Table 1. Design of Two Level Factorial Experiment

EXPERIMENT WORKSHEET

NAME OF EXPERIMENT	: TOOLS3
NUMBER OF FACTORS	: 10
RESOLUTION	: 3
NUMBER OF RUNS	: 32
DESIGN FRACTION	: 1/32
BLOCKING	: NONE
NUMBER OF CENTER POINTS PER BLOCK	: 0
NUMBER OF REPLICATIONS WITHIN BLOCK	: -
NUMBER OF BLOCK REPLICATIONS	: -
FACTOR CODES	: ABCDEVWXYZ
SPECIFIED ESTIMABLE INTERACTIONS	: AV;BW;CX;DY;EZ

RUN	A	B	C	D	E	V	W	X	Y	Z	CYCLE TIMES	EST. VARIANCES
1	-	-	-	+	+	-	+	+	-	+	373.94	2.61
2	+	-	+	-	+	-	+	-	+	-	378.33	12.70
3	-	+	-	+	-	-	+	-	+	+	370.94	5.30
4	+	+	-	-	+	+	+	+	+	-	377.09	6.08
5	-	+	+	+	+	+	+	-	-	+	363.11	6.33
6	+	+	+	+	+	+	-	+	-	-	364.31	4.74
7	-	-	+	+	-	+	+	+	+	+	378.25	1.50
8	-	-	-	-	-	+	-	+	+	-	384.40	6.90
9	-	+	-	-	+	+	-	-	-	-	373.22	6.67
10	+	+	-	-	+	+	+	+	-	+	377.76	4.69
11	-	+	+	-	-	-	-	-	+	-	377.58	7.48
12	-	-	-	-	-	+	-	+	-	+	387.75	5.01
13	+	-	-	-	-	+	+	-	-	-	385.08	6.16
14	+	-	+	+	-	+	-	-	+	-	367.26	4.53
15	-	+	+	-	-	-	-	-	-	+	384.32	6.76
16	+	+	+	-	-	-	+	+	+	+	376.60	4.39
17	+	-	-	+	+	-	-	-	-	-	370.00	7.86
18	+	+	+	+	+	+	-	+	+	+	361.67	2.53
19	-	-	-	+	+	-	+	+	+	-	372.73	3.80
20	+	-	-	+	+	-	-	-	+	+	368.92	3.84
21	+	+	-	+	-	-	-	+	-	+	371.95	3.11
22	-	-	+	-	+	-	-	+	-	-	382.67	6.38
23	-	+	+	+	+	+	+	-	+	-	361.25	5.65
24	-	+	-	+	-	-	+	-	-	-	367.22	3.17
25	+	-	+	-	+	-	+	-	-	+	373.44	12.22
26	+	+	+	-	-	-	+	+	-	-	371.35	13.07
27	+	-	+	+	-	+	-	-	-	+	373.96	5.51
28	-	+	-	-	+	+	-	-	+	+	375.54	5.99
29	+	-	-	-	-	+	+	-	+	+	387.98	3.39
30	-	-	+	-	+	-	-	+	+	+	379.30	10.18
31	-	-	+	+	-	+	+	+	-	-	368.96	2.28
32	+	+	-	+	-	-	-	+	+	-	367.69	3.26

model is shown in Table 2. Notice that 8 of the 10 main effects and 3 of the 5 interactions are confounded with 2 factor interactions. To investigate the cost associated with removing this weakness we also requested a resolution 4 design with AV, ... , EZ not confounded with any main effects, one another, or with any other 2 factor interactions. In this design the effects of the proposed model were not confounded with any two factor interactions outside the model. In a resolution 4 design no main effect is confounded with any second order interaction. However this design required 64 runs and the advantages did not seem worth the extra cost of doubling the number of runs.

Table 2. Alias Structure

EFFECT	ALIASES
A	BX DW
B	AX EV
C	DV
D	AW CV
E	BV
V	BE CD
W	AD
X	AB
Y	none
Z	none
AV	CW EX
BW	DX
CX	EW YZ
DY	none
EZ	WY

In this design selection phase the availability of flexible, interactive software was essential. For example, the ability to specify the estimability of a set of interactions subject to a resolution constraint with a given number of runs was critical to arriving at an effective design. Each design took just a few minutes to generate, including the time to type in the input data (the information at the top of Table 1).

3. THE ANALYSIS OF THE INPUT DATA

Thirty-two runs were made with the parameter settings shown in Table 1. Each run was slightly longer than four thousand observations for the product studied. This corresponds to an operating period of approximately three years. The output series corresponding to the first run is shown in Figure 1. It is representative of all thirty-two output series. All the output series were first viewed to make sure the model was in steady state and to assess the number of observations to be dropped to eliminate the initial transient. In all cases steady state was reached during the first 400 observations so 400 observations were dropped from the beginning of each series to generate the series used to obtain the responses. Specifically, each response is the mean of the truncated series.

In this environment it is desirable to obtain internal estimates of the variance of the responses, independent of the model fitting process. This is important for two reasons. First, one can check that the estimates are consistent with the assumption of a constant error variance. Second, one can obtain a pure error estimate which is independent of any model assumptions and which can be compared with the error estimate obtained from the residuals to the fitted model. To

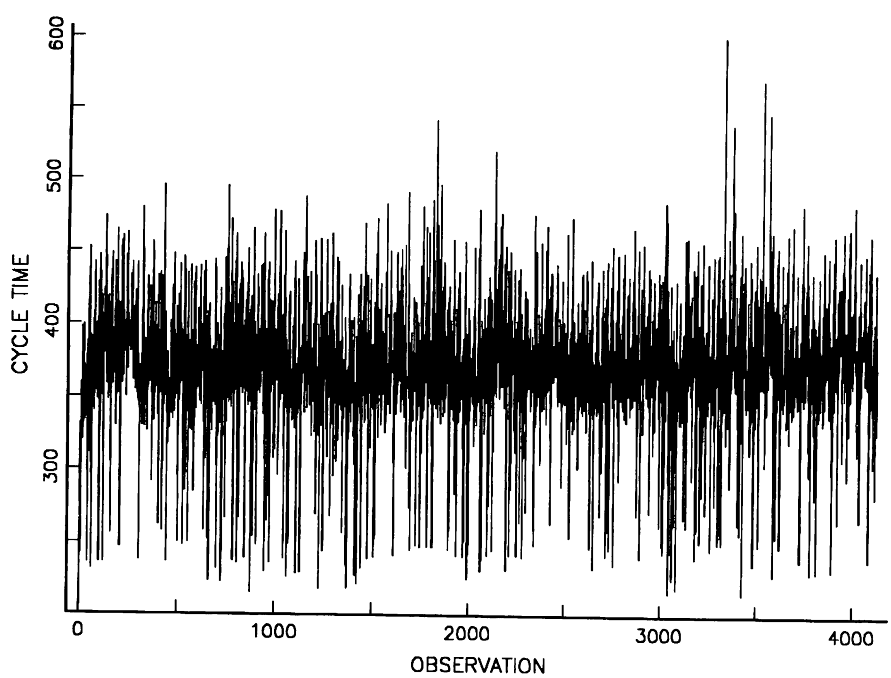

Figure 1. Cycle Time versus Observation Number for Run 1

this end the method of batch means was used to obtain approximately uncorrelated replications at each experimental point.

An examination of the thirty-two correlation functions showed that the correlation dropped abruptly but remained significantly different from zero out to about 60 to 80 lags. In Figure 2 the sample correlation function for the first run is plotted. The dotted limits on the tail should contain the function 95% of the time assuming the true correlation function is zero for lags greater than 60 [Box and Jenkins 1976 Sections 2.1.6 and 6.2.2.]. When we batched the data into batches of size 100 they passed this correlation function test for white noise so, applying a margin of safety, we generated ten batches of 360 observations for each series and assumed the batch means to be uncorrelated. Hence for each of the thirty-two runs we obtained, as an estimate of the variance of the mean, the sample variance of the ten batches divided by 10. These variance estimates are given in Table 1.

The statistical tests and procedures underlying the design of experiments assume that all the responses have a common variance. To test this assumption Bartlett's test was applied to the thirty-two sample variances. They were found to be consistent with the assumption of a constant variance. As an additional graphical test the distribution of the sample variances was examined. Let s_i^2 be the sample variance for the *ith* run and let s^2 be the average of these sample variances then $9s_i^2/s^2$ should be distributed approximately as a chi-squared random variable with nine degrees of freedom. The sample distribution was compared with the theoretical distribution both graphically (Figure 3) and using goodness-of-fit tests. The sample distribution was consistent with the assumption of a common variance. The first plot shows the histogram overlaid with theoretical density. The second plot shows the empirical cumulative distribution function (c.d.f.) overlaid with the theoretical c.d.f.. The third plot shows the two c.d.f.'s on probability paper such that the theoretical c.d.f. plots as a straight line. After fitting the model additional graphical tests of the common variance assumption were made and will be discussed below. For a good discussion of this issue see Hald [1952], Section 11.6. The final internal estimate of the error variance was the average of the sample variances, $s^2 = 5.75$.

4. THE ANALYSIS OF THE EXPERIMENT

The thirty-two measurements of the response, the dependent variable, are given in the column labeled "CYCLE TIMES" in Table 1. As described above, they are the means of the thirty-two series after removal of the effects of the initial transient. Each of the truncated series was approximately 3600 observations long. In addition to this data the independently estimated error variance, as described earlier, was 5.75. The GRAFSTAT system provides a number of tabular and graphical outputs for the analysis of the experiment. These include tables of estimates with t-confidence intervals, an analysis of variance table, a probability plot of effects and a number of plots and tables for analyzing the residuals.

In the discussion below three questions are addressed. First, is the the 16 variable model described in Section 1 justified? That is, is it reasonable to neglect all higher order interactions with the exception of AV, BW, CX, DY, and EZ? Second, if so, do the results indicate a model with even fewer parameters? That is, are any of the 15 effects of the model negligible? Third, how does one summarize the meaning of the final model?

Table 3 contains the estimates of the fifteen effects along with 95% confidence intervals. If the confidence intervals do not contain zero then the effects are significant at the 0.05 level. Care has to be used in viewing such a table because, since there are fifteen such effects, a few are very likely to be significant by pure chance. The confidence statement holds strictly only for an effect chosen *a priori*, not for one selected from amongst the group on the basis of the confidence intervals. This matter will be discussed again later when the probability plot of the effects is examined. In Table 3 the effects A, B, C, D, E are significant at the 0.05 level. Hence all the main effects due to the number of tools are justified in the model. Table 3 shows also that none of the main effects due to the operator to tool ratio are significant except for Z and that effect is positive! Furthermore, only two of the

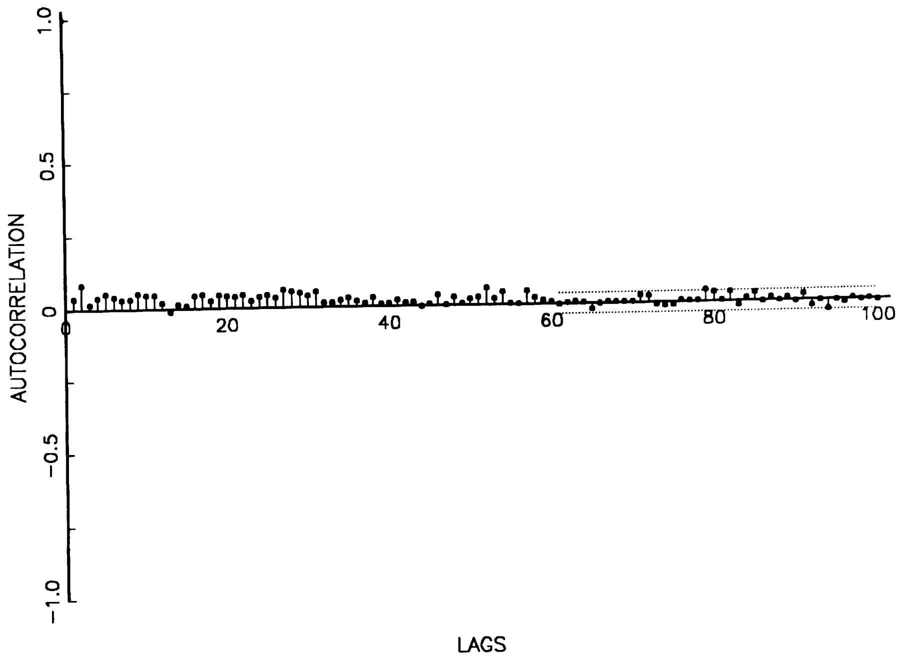

Figure 2. Sample Correlation Function for Run 1

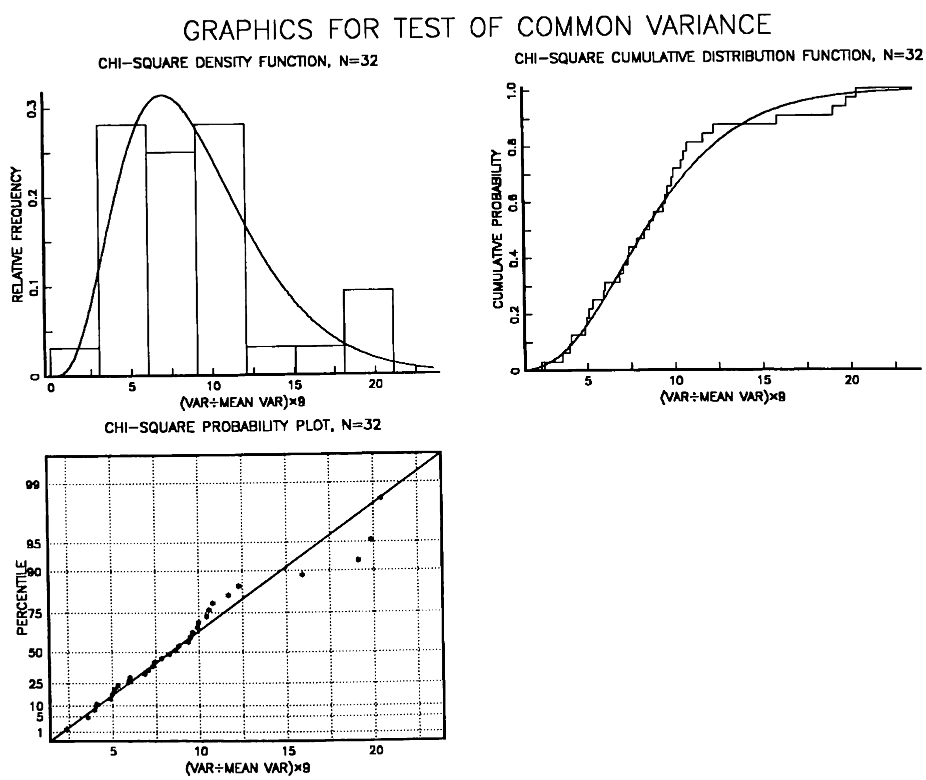

Figure 3. Sample and Theoretical Distributions for the Estimated Variances

Table 3. Table of Coefficient Values and Confidence Intervals

```
32 OBSERVATIONS    R-SQUARED     = 0.95762    STANDARD ERROR = 2.0938
16 VARIABLES       ADJ R-SQUARED = 0.91789
```

EFFECT	ESTIMATE	STD ERR	T STAT	SIG LEVEL	0.95 CONFIDENCE LIMITS LOWER	UPPER
MEAN	374.21	0.37013	1011	8.3267E-17	373.42	374.99
A	-1.7375	0.74026	-2.3472	3.2114E-2	-3.3069	-0.16816
B	-5.7105	0.74026	-7.7142	8.8619E-7	-7.2798	-4.1411
C	-3.1171	0.74026	-4.2108	6.6340E-4	-4.6864	-1.5477
D	-10.641	0.74026	-14.375	1.4438E-10	-12.211	-9.072
E	-4.2504	0.74026	-5.7418	3.0308E-5	-5.8198	-2.6811
V	0.037602	0.74026	0.050795	9.6012E-1	-1.5318	1.607
W	-0.40675	0.74026	-0.54946	5.9027E-1	-1.9761	1.1626
X	1.142	0.74026	1.5427	1.4245E-1	-0.42736	2.7114
Y	-0.22023	0.74026	-0.2975	7.6991E-1	-1.7896	1.3491
Z	2.2677	0.74026	3.0634	7.4269E-3	0.69839	3.8371
AV	2.0652	0.74026	2.7898	1.3113E-2	0.49585	3.6346
BW	-0.96322	0.74026	-1.3012	2.1162E-1	-2.5326	0.60614
CX	-0.65953	0.74026	-0.89094	3.8616E-1	-2.2289	0.90983
DY	-0.37269	0.74026	-0.50346	6.2150E-1	-1.942	1.1967
EZ	-3.0077	0.74026	-4.0631	9.0393E-4	-4.5771	-1.4384

interactions, AV and EZ, are significant. The analysis of variance table gave the estimate of the error variance obtained from the sixteen degrees of freedom orthogonal to the model (i.e. the degrees of freedom for error) as 4.38. This is consistent with the independent assessment of the error variance of 5.75 obtained from the analysis of the estimated variances of the batch means thus providing confirmation for the legitimacy of the sixteen parameter model.

Next consider Figure 4, a normal probability plot of the effects. This is a plot of the effects on normal probability paper. The fifteen effects of the model are labeled. The sixteen effects orthogonal to the model are not. If none of the effects were significant, these 31 values would have a normal distribution with mean zero and hence lie on a straight line through the point (0,50). Notice that the effects A, B, C, D, E, Z, AV and EZ are inconsistent with this. This confirms the conclusions drawn from Table 3. Furthermore, viewing effects in such a plot provides reasonable guidance in the matter of judging the significance of selected effects because they are viewed in the context of the expected normal distribution of insignificant effects. As an additional test Figure 5 shows a normal probability plot of the 23 effects orthogonal to the suggested nine parameter model. The X axis scale is the same as that of Figure 4 so an easy comparison can be made. Notice that the 23 points do lie on a straight line through (0,50). Hence the 16 error effects and the 7 effects V, W, X, Y, BW, CX and DY are approximately normally distributed with a variance consistent with the internal estimate of 5.75 This is additional confirmation of a nine parameter model and hence of the sixteen parameter model.

Figure 6 contains a number of plots diagnosing the distribution of the residuals from the 9 parameter model. These plots and the tests associated with them show the residuals to be approximately normally distributed. The plot of the resi-

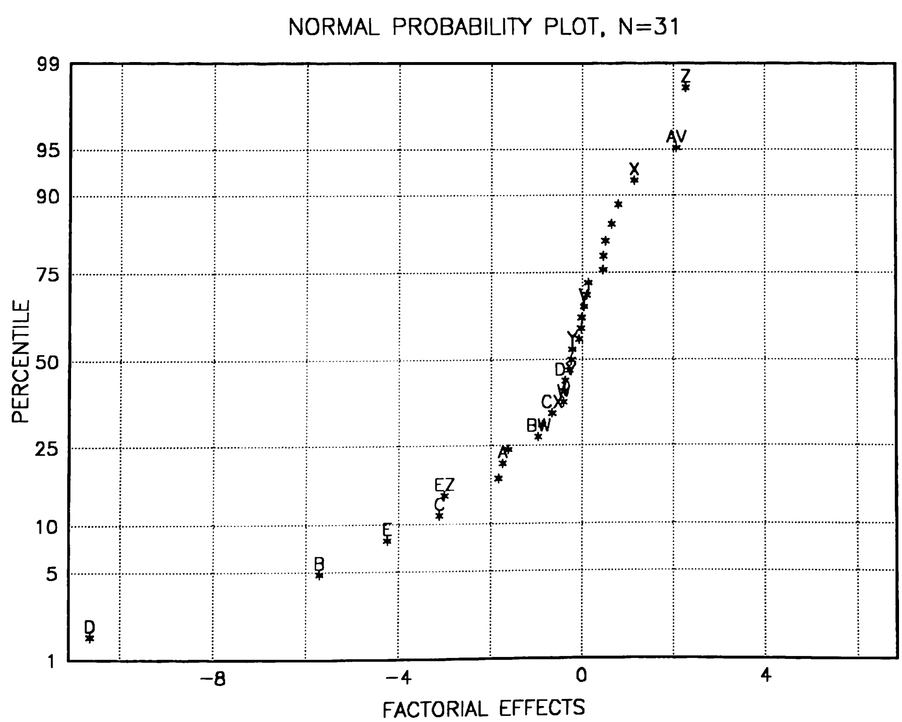

Figure 4. Normal Probability Plot for All 31 Effects

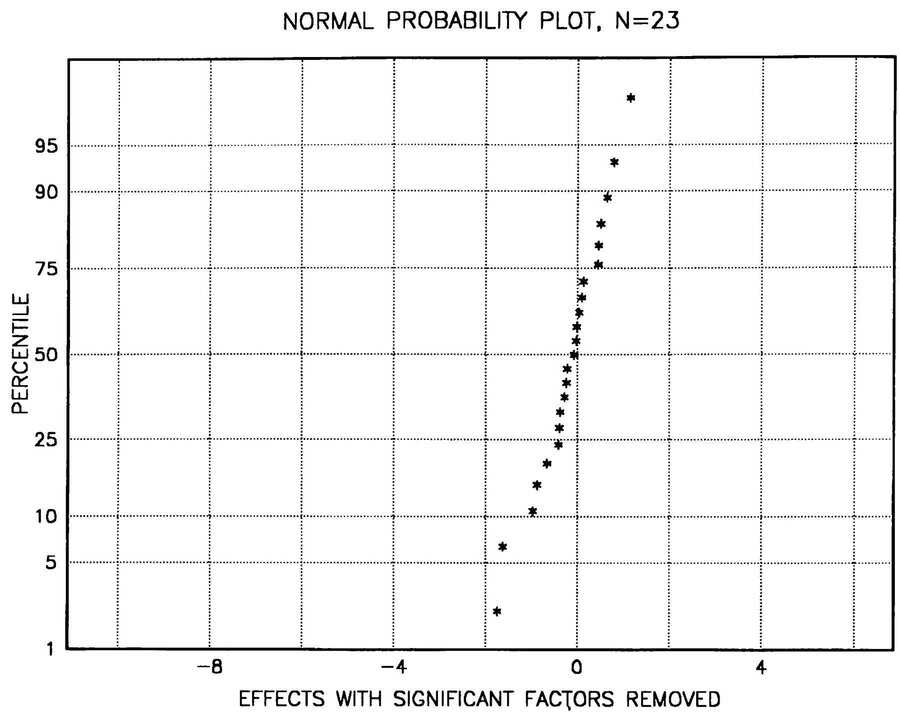

Figure 5. Normal Probability Plot for the 23 Non-model Effects

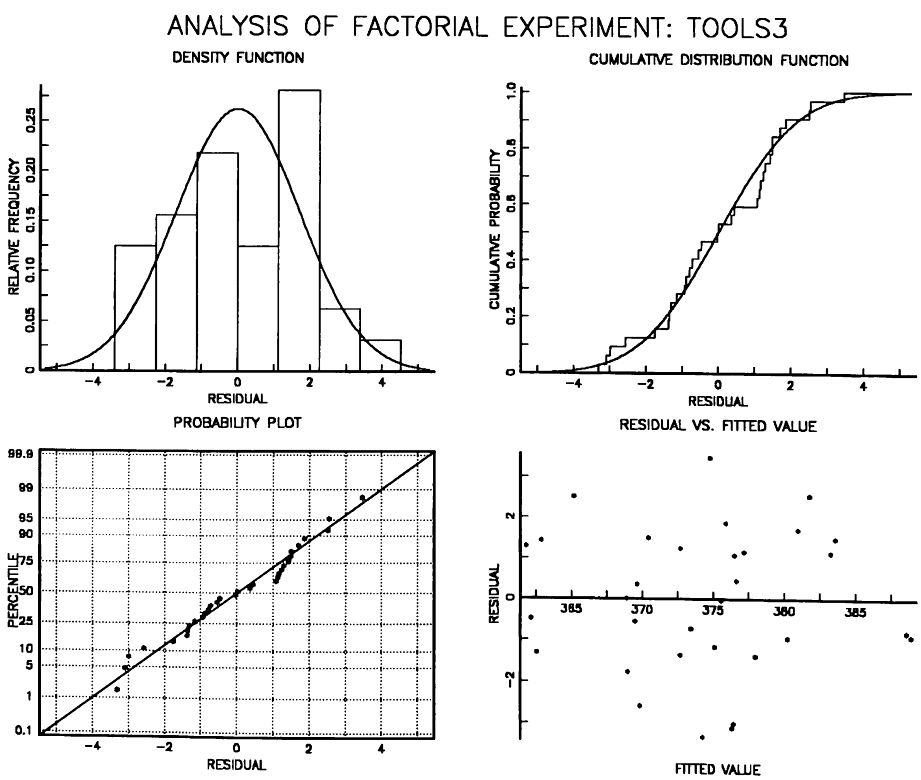

Figure 6. Plots of Residuals from the 9 Parameter Model

duals against the fitted values indicates that the variance of these residuals does not depend on the value of the response, confirming the results of the common variance tests described earlier. Finally, in Figure 7 the estimated variances at each experimental point (the last column of Table 3) are plotted against the fitted values. Again there appears to be no dependence of the variance on the value of the response.

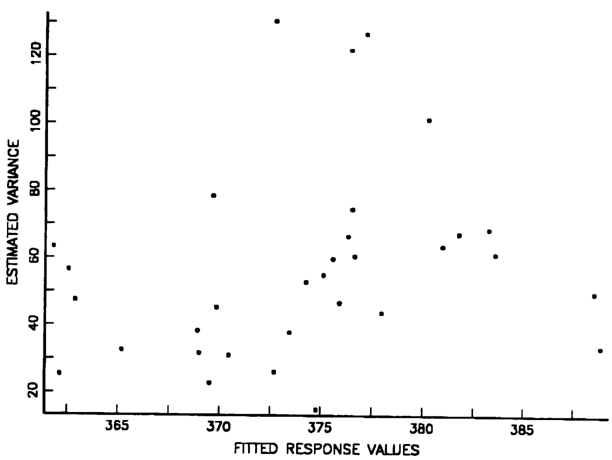

Figure 7. Estimated Variances versus Fitted Values

Thus the final model is of the form

$$R(\mu,\ x_1, \dots,\ x_5,\ y_1, \dots, y_5)$$
$$= \mu + \sum_{i=1}^{5} \alpha_i x_i + \beta_5 y_5 + \gamma_{11} x_1 y_1 + \gamma_{55} x_5 y_5 .$$

The situation with tool groups B, C and D is clear. Adding tools reduces the cycle time by the estimated amounts of 5.71, 3.12 and 10.64 units respectively. Changing the operator to tool ratio has no significant effect. The situation with tool groups A and E is more complex because of the presence of other significant effects. The estimated model can best be understood by considering the 2 by 2 effect arrays. For tool group A, where in addition to the main effect there is the interaction AV, the array is

```
A  +   -1.90    .16
   -    1.90   -.16
         -       +
         V
```

Because of the positive sign of the AV interaction, the benefit of adding tools (row 1) is eliminated if the operator to tool ratio is changed to 1:1. Again as with tool groups B, C, and D, to reduce cycle time only the number of tools would be increased, not the operator to tool ratio. For tool group E the situation is even more complex because of the significance of the two main effects, E and Z, and the two factor interaction EZ. In this case the array is

```
E  +   -1.76   -2.50
   -   - .51    4.76
         -       +
         Z
```

Here to change the operator to tool ratio to 1:1 (column 2) without the addition of tools would result in an *increase* in the cycle time. Further to change it to 1:1 with the addition of tools results in only a marginal reduction in cycle time over the addition of tools alone (element 1,2 vs. element 1,1). These results for tool group E are puzzling and need to be subject to further investigation.

5. SUMMARY

The analysis of the experiment indicates that the model proposed in section 1 is a reasonable one. A review of ancillary output data from each run corroborates the results of the experimental analysis, namely that tool effects predominate. It reveals that the main effects for the tools are ordered according to utilization. At both levels for the factor, tool group D, which has the largest tool effect, has less than 10% idle time. The rest of the time is spent for processing jobs, setup, preventative maintenance, and failure. Tool group B, which has the second largest tool effect, is the next highly utilized, and A, C and E all have approximately the same utilizations. The fact that the tool groups spend so little time idle indicates that operators are not the constraining resource.

The flexible, interactive software for designing the experiments provided the capability of investigating several designs. In just a few minutes we were able to improve the design, going from 16 runs to 32 runs with specified estimable interactions, and rejecting a 64 run design as not giving enough benefit to make it worth the cost of doing the extra runs. A full factorial design would have required 1024 runs to estimate the effect of the ten factors, clearly out of the question.

Thus with a highly efficient experiment we were able to determine that the cycle time was sensitive to the number of tools and insensitive to the operator to tool ratio. We were able to order the tool sensitivity in a way consistent with tool utilization. We determined that, in only two cases, was there a significant interaction between number of tools and the operator to tool ratio. Hence the experiment confirmed a simple 16 parameter model and suggested an even simpler 9 parameter model.

REFERENCES

Box, G.E.P., W.G. Hunter, and J.S. Hunter (1978), *Statistics for Experimenters*, John Wiley & Sons.

Box, G.E.P. and G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, Holden Day.

Hald, A. (1952), *Statistical Theory with Engineering Applications* John Wiley & Sons.

Hood, S.J., A.E. Amamoto, and A.T. Vandenberge (1989), "A Modular Structure for A Highly Detailed Model of Semiconductor Manufacturing," In *Proceedings of the 1989 Winter Simulation Conference*, E.A. MacNair, K.J. Musselman, and P. Heidelberger, Eds. IEEE, Piscataway, NJ, 811-817.

Lane, T. and P.D. Welch (1987), "The Integration of a Menu-Oriented Graphical Statistical System with its Underlying General Purpose Language" *Computer Science and Statistics: Proceedings of the 19th Symposium of the Interface*, American Statistical Association, 267-273.