

# Designing Experiments for the Modern Micro Industries

Although the sparsity of effects principle is still good advice, we need to be careful of emphasizing it too much — to the point that it becomes an assumed practice rather than a useful rule of thumb.

**PHILLIP H. WILLIAMS**  
FREESCALE SEMICONDUCTOR, INC.

**H**igher-order interactions are less likely to be significant than lower-order interactions. This is the essence of the “sparsity of effects” principle that is continuously emphasized in statistics and design-of-experiments (DoE) training. Historically, in theory and in practice, this rule of thumb is well substantiated. But as with most rules of thumb, we need to consider its origin, and its applicability in the face of new and evolving technologies.

The DoE class manual of a large technology-driven company typically reads — *If you have five factors and need information on all of the factors, never do the full factorial since the  $2^{(5-1)}$  design is a resolution-five design.*

Generally, this is good advice from a resources standpoint. But, if there is one thing that both Murphy and our many technological endeavors have taught us, it is “Never say never.” The above quote is more than just a non-optimal choice of words. Rather, it reflects a mentality that is widespread in experimental design education and practice.

Higher-order interactions (typically third-order or higher) are rarely considered in the examples given to students learning about the statistical DoE. Mainly, this is done for purposes of simplicity — but it is often argued as being justified because of the sparsity of effects principle. In any event, students walk away with the idea firmly planted in their minds that higher-order interactions are not important.

Is this such a bad thing? Yes it is, if it causes us to remove all experiments from their technological contexts, and then design and analyze all of them exactly the same — a sort of experimental one-size-fits-all mentality.

The design should fit the experiment, not vice versa. The design should be consistent with the ease of experimentation, the complexity of the process being investigated, and the available resources. The uniqueness of the technology, and the uniqueness of the particular process or phenomenon being investigated should always be considered.

Let’s take an informal look at the origin of the sparsity of effects principle. “Modern” industrial statistical methods were essentially derived during the 1920s through the 1950s. The sparsity of effects principle rose out of this era, an era of fundamental “macro” advancements in agriculture, food processing, aerospace and automobile manufacture, weapons, civil construction and medicine.

The later half of the 20th century, however, gave rise to the “micro” industries — microelectronics, complex pharmaceuticals, genetic engineering, complex medical testing, nanotechnology, etc. These modern “micro” industries are most often characterized by very complex transient/batch processes that utilize complex materials and energy sources to achieve the desired result. This dramatically increases the opportunity for many factors to interact in a complex fashion.

In addition, these “micro” products tend to have a multitude of traits or characteristics in which we are interested. Rather than optimizing one or two product characteristics, as is typical with the “macro” products, there tends to be a multitude of characteristics that need to be pseudo-optimized in parallel to achieve the desired goal. It stands to reason that the more we have of these “simultaneous” characteristics, the more likely interactions — including high-order interactions — will play a significant role.

So can’t we just experiment sequentially to obtain all the desired information, that is, perform one fraction of the experiment now, another fraction later, and so forth? Isn’t this the spirit of the sparsity of effects principle? Yes, it is the *spirit*, but unfortunately it is oftentimes not the *practice*. Instead, the experimentation is halted prematurely because it is believed that the higher-order interactions are not worth investigating.

In addition, each time a fraction of the experiment is performed, a blocking factor should be included. This causes information on a higher-order interaction to be lost. It also adds to the complexity of the analysis. Finally, if we highly

suspect that higher-order interactions will play a significant role, then we may be wasting valuable time and resources by the sequential approach.

So, should we continue to drill students and practitioners on the sparsity of effects principle with the same vigor as in the past? If we do, we are not matching the tools to the era in which we live. And using our tools in an inappropriate fashion may lead to inappropriate conclusions — and in the long run inefficient experimentation.

Through my many years of experience as a plasma-enhanced CVD engineer, I learned early that “once you strike a plasma, you may as well throw the chemistry book away.” Anyone who has endeavored to characterize and understand a plasma-assisted process (and the complexities of what it leaves behind) can empathize with this rather facetious statement. What it really means is that we can no longer depend on traditional concepts (such as that only lower-order interactions are significant) to always explain the results.

The best way to illustrate the above precepts is by way of an example. The example herein is based on real data taken from a real process (not a computer simulation). Although it is relatively simple, it is representative of the sort of complex and multi-factorial processes that are encountered frequently in the microelectronics industry.

### Designing the experiment

In this example, solder bumps are being electroplated onto finished wafer substrates in preparation for flip-chip packaging. Five of the tool inputs are known to impact the plated bump characteristics. Three of the inputs — A, B and C — are continuous, and the remaining two inputs (D and E) are discrete (yes/no or on/off). A two-level five-factor full factorial ( $2^5$ ) experiment was designed and performed in order to characterize the impacts of these factors on the solder bump formation. Due to the relative ease of experimentation, running a fractional factorial was not considered.

An important response — the one evaluated in this example — is the bump diameter,  $C_p$ . The  $C_p$  is the “process capability potential” corresponding to each of the factor combinations, and is defined as

$$C_p = \frac{\text{bump-diameter tolerance}}{6 \times \text{bump-diameter standard deviation}}$$

For each wafer corresponding to a run or factor combination, an automated metrology tool evaluated the effective diameters of thousands of bumps. The standard deviation of

Table 1.  $2^5$  full factorial design and responses.\*

Std	Run	A	B	C	D	E	$C_p$
Order	Order						
1	23	-1	-1	-1	-1	-1	3.1396
2	27	1	-1	-1	-1	-1	1.7093
3	11	-1	1	-1	-1	-1	2.7775
4	20	1	1	-1	-1	-1	1.9263
5	16	-1	-1	1	-1	-1	4.0780
6	7	1	-1	1	-1	-1	2.4061
7	4	-1	1	1	-1	-1	3.4840
8	19	1	1	1	-1	-1	1.9339
9	32	-1	-1	-1	1	-1	2.7571
10	22	1	-1	-1	1	-1	1.9450
11	30	-1	1	-1	1	-1	1.8837
12	13	1	1	-1	1	-1	1.7790
13	8	-1	-1	1	1	-1	3.2085
14	15	1	-1	1	1	-1	2.2596
15	10	-1	1	1	1	-1	2.1334
16	14	1	1	1	1	-1	1.7632
17	29	-1	-1	-1	-1	1	6.1258
18	1	1	-1	-1	-1	1	4.9905
19	9	-1	1	-1	-1	1	2.3759
20	25	1	1	-1	-1	1	2.0300
21	5	-1	-1	1	-1	1	6.2670
24	17	1	1	1	-1	1	2.0021
22	2	1	-1	1	-1	1	4.6765
23	24	-1	1	1	-1	1	2.3445
25	12	-1	-1	-1	1	1	5.1000
26	31	1	-1	-1	1	1	1.7881
27	28	-1	1	-1	1	1	2.3129
28	3	1	1	-1	1	1	2.1038
29	21	-1	-1	1	1	1	5.1803
30	18	1	-1	1	1	1	2.2081
31	26	-1	1	1	1	1	2.3641
32	6	1	1	1	1	1	2.1552

\*The shaded runs correspond to the principle half-fraction.

each bump population was then calculated and entered into the above definition to generate the  $C_p$  response value. The  $2^5$  full-factorial design is given in Table 1, along with the actual response values.

### Devising an analysis strategy

The data will be analyzed as a full factorial using all of the data, and then re-analyzed as a half fraction of the full factorial. For the half fraction, the data subset corresponding to the principle fraction will be used (see the shaded runs in Table 1). The results of these two analyses will then be directly compared.

The half-fraction analysis will represent the information

# Process Design

Table 2. Coefficients and their p-values for the reduced full-factorial model.

Factorial fit: <i>C<sub>p</sub></i> vs. A, B, C, D, E					
Term	Effect	Coefficient	SE	T	P
			Coefficient		
Constant		2.9128	0.04303	67.70	0.000
A	-1.1160	-0.5580	0.04303	-12.97	0.000
B	-1.4044	-0.7022	0.04303	-16.32	0.000
C	0.2325	0.1163	0.04303	2.70	0.016
D	-0.7078	-0.3539	0.04303	-8.23	0.000
E	0.9275	0.4638	0.04303	10.78	0.000
AB	0.6182	0.3091	0.04303	7.18	0.000
AD	-0.0013	-0.0006	0.04303	-0.01	0.988
AE	-0.1486	-0.0743	0.04303	-1.73	0.105
BD	0.4105	0.2052	0.04303	4.77	0.000
BE	-0.9266	-0.4633	0.04303	-10.77	0.000
CE	-0.1862	-0.0931	0.04303	-2.16	0.047
DE	-0.2422	-0.1211	0.04303	-2.81	0.013
ABD	0.2759	0.1379	0.04303	3.21	0.006
ABE	0.3698	0.1849	0.04303	4.30	0.001
ADE	-0.4097	-0.2049	0.04303	-4.76	0.000
BDE	0.5854	0.2927	0.04303	6.80	0.000

S = 0.243394      R-Square = 98.41%      Adjusted R-Square = 96.72%

Analysis of Variance for <i>C<sub>p</sub></i> (coded units)						
Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	5	37.0644	37.0644	7.41288	125.13	0.000
2-Way Interactions	7	12.1964	12.1964	1.74234	29.41	0.000
3-Way Interactions	4	5.7873	5.7873	1.44682	24.42	0.000
Residual Error	15	0.8886	0.8886	0.05924		
Total	31	55.9367				

Note: SE Coefficient = standard error of the coefficient; T = T-statistic; P = the p-value, or probability value; DF = degrees of freedom; Seq SS = sequential sum of squares; Adj SS = adjusted sum of squares; Adj MS = adjusted mean square; F = the F statistic (a ratio of variances or mean squares).

this, however, we will still achieve a higher experimental resolution than for the half-fraction analysis: All of the main effects, two-factor interactions, and three-factor interactions can be estimated free and clear of one another. This is contrasted by the half-fraction analysis, in which we cannot estimate the second-order interactions free and clear of the third-order interactions. That is, an apparently significant second-order interaction may, in actuality, be a significant third-order interaction. This is not likely to be the case according to the sparsity of effects principle. In fact, it is oftentimes dismissed both philosophically — and due to the lack of data. This dismissal is convenient, since oftentimes only the half-fraction is performed with no intention to run the other half.

In order to produce an experimental error for the half-fraction analysis, we need to “trick” the analysis software (MINITAB v14). Recall that in the full-factorial analysis, we are using the fourth- and fifth-order terms to estimate the experimental error. But in the half-fraction analysis, these terms are not available.

In theory, the experimental error associated with the fractional factorial will be the same as for the full factorial. Since the experimental error for the full factorial was assigned to the fourth- and fifth-order interactions, a mean square error (MSE) was generated. This error was then artificially “inserted” into the half-fraction analysis by including two false centerpoints (determined by trial and error) that resulted in the same MSE. This then allowed us to determine the significances associated with the fractional factorial effects. Once this was done, the model could be reduced, the false centerpoints removed, and the MSE recalculated based on the eliminated model terms — the same as we did for the full factorial analysis.

### Full-factorial results

The analysis of the full-factorial matrix shown in Table 1 produced the Pareto chart of all effects shown in Figure 1. Note that with the exception of one 4-factor interaction ABDE, all of the 4-factor and 5-factor interactions are in the lower half of the Pareto chart (so our method that we will use to estimate the error is reasonable). But also note that there are many 3-factor and 2-factor interactions in the upper half of the Pareto. In fact, there are several third-order effects that are more significant than one of the main effects (C).

Removing the fourth- and fifth-order interactions from the model, as previously discussed, allows us to estimate the experimental error. Thus, we can now obtain an estimate of the absolute significances, or p-values, associated with the remaining model terms. The Pareto resulting from this analysis is shown in Figure 2. The vertical reference line demarks the significant effects from the non-significant effects, using

obtained and the conclusions drawn as if the experimenters considered only the lower-order interactions. By comparing these conclusions to those obtained from the full analysis, we will then be able to discuss the ramifications of what was missed by only performing the half-fraction experiment.

An experiment that is perfect in its design and/or execution is truly rare. This experiment was not an exception, in that centerpoints were not performed (because there are two categorical factors, a minimum of eight centerpoints would have been required for a balanced set). So, in order to estimate the experimental error, it was decided to consider the fourth- and fifth-order interactions as pure error or experimental noise.

Ironically, we are calling upon the sparsity of effects principle to help us evaluate its own applicability! Even in doing

**Table 3. Coefficients and their p-values for the fractional-factorial model (reduced).**

**Factorial fit: Cp vs. A, B, C, D, E**

Term	Effect	Coefficient	SE	T	P
Constant		2.8776	0.05597	51.41	0.000
A	-1.1722	-0.5861	0.05597	-10.47	0.002
B	-1.3885	-0.6943	0.05597	-12.40	0.001
C	0.4352	0.2176	0.05597	3.89	0.030
D	-0.6637	-0.3319	0.05597	-5.93	0.010
E	0.8982	0.4491	0.05597	8.02	0.004
AB	0.7547	0.3773	0.05597	6.74	0.007
AC	0.4945	0.2472	0.05597	4.42	0.022
BC	-0.5183	-0.2592	0.05597	-4.63	0.019
BD	0.4874	0.2437	0.05597	4.35	0.022
BE	-0.8435	-0.4217	0.05597	-7.53	0.005
CD	0.3376	0.1688	0.05597	3.02	0.057
DE	-0.2713	-0.1357	0.05597	-2.42	0.094

S = 0.223894      R-Square = 99.46%      Adjusted R-square = 97.31%

**Analysis of Variance for Cp (coded units)**

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	5	18.9555	18.9555	3.79110	75.63	0.002
2-Way Interactions	7	8.8774	8.8774	1.26821	25.30	0.011
Residual Error	3	0.1504	0.1504	0.05013		
Total	15	27.9833				

Note: See Table 2 for definitions of abbreviations.

the fourth- and fifth-order terms to estimate the experimental error and assuming an alpha-risk of 0.10. Once again, it can be noted that there are several significant third-order effects. Residual diagnostics indicate the model assumptions of constant variance, independence, and normality are reasonably valid.

The MINITAB-generated coefficients and p-values, as well as the R-square and adjusted R-square of the reduced full-factorial model are given in Table 2. The non-significant terms (*i.e.*, terms with a p-value greater than 0.10) have been eliminated. Note that the second-order term AD was kept even though it has a high p-value. This is because it is contained in the significant higher-order terms ADE and ABD. The adjusted R-square of 96.7% indicates that the model explains the variation in the response to a high degree.

### Fractional-factorial results

Only data from the principle half-fraction of the experiment was reanalyzed. This will represent the information obtained and the conclusions drawn as if the experimenters considered only the lower-order interactions.

Also recall that no centerpoints were run. Therefore, in order to distinguish the non-significant terms, two false centerpoints were included in the data set, as previously noted. These centerpoints were chosen (by trial and error) to result in the same initial estimate of experimental error, or MSE, as was determined in the full-factorial analysis by the elimination of the fourth-

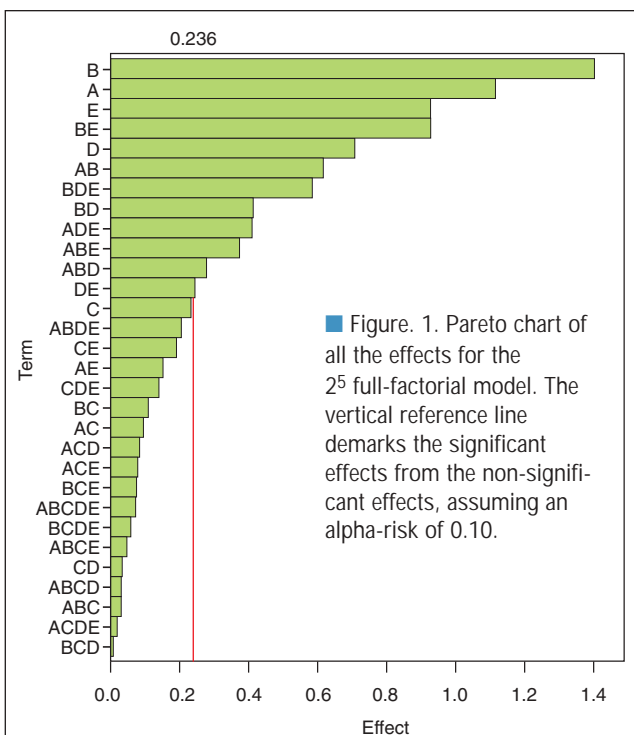


Figure 1. Pareto chart of all the effects for the 2<sup>5</sup> full-factorial model. The vertical reference line demarks the significant effects from the non-significant effects, assuming an alpha-risk of 0.10.

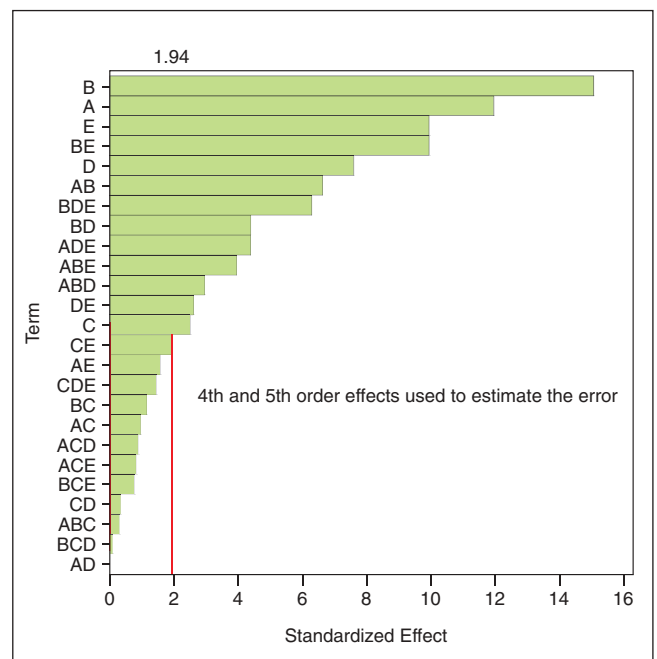


Figure 2. Pareto chart of all main, second- and third-order effects for the 2<sup>5</sup> full factorial model.

# Process Design

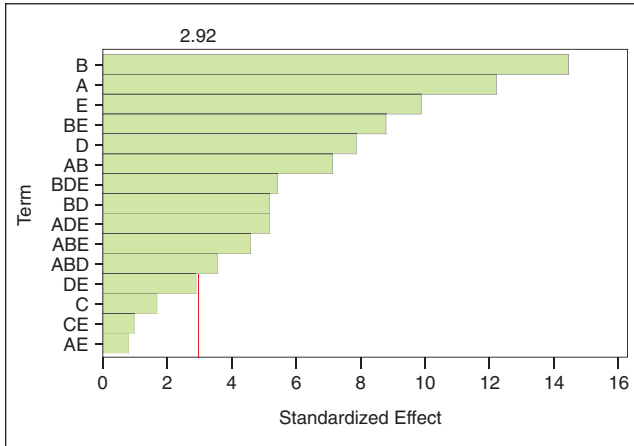


Figure 3. Pareto chart for the fractional-factorial model effects (unreduced).

and fifth-order interaction terms. Again, the experimental runs and their responses are the shaded runs in Table 1.

Figure 3 shows the Pareto chart for the model effects. From this, it can be seen that only three effects are not significant (the DE interaction, with  $p = 0.10$ , will be considered significant). Again, residual diagnostics indicated that the assumptions of independence, constant variance and normality are reasonably valid. With the non-significant model terms determined, both these and the false centerpoints are now removed from the analysis, and the  $p$ -values recalculated. The experimental error is now estimated using the non-significant model terms, the same as was done for the full-factorial analysis. Table 3 shows the model coefficients and the  $p$ -values corresponding to the reduced model.

## Comparing full- vs. fractional-factorial results

Table 4 is a combined summary of the full-factorial and fractional-factorial analyses. The first column lists all of the effects, or model terms, up through the third order. The second and third columns list the significant model coefficients for each of the analyses. These are simply copied from Tables 2 and 3. The fourth column lists the three-factor alias corresponding to each of the fractional factorial two-factor interactions.

Of particular interest in this table are the significant coefficients that were identified using the fractional-factorial analysis, but were not identified using the full-factorial analysis. These are the highlighted interactions AC, BC and CD. These were identified as being significant, when in fact they are not. Instead, it is the indicated three-factor interactions — BDE, ADE, and ABE, respectively — that are the *real* significant terms. Additionally, the fractional analysis failed to identify the significant three-factor interaction ABD.

Figure 4 is a graphical comparison of the results. From this, it is clear that the four significant three-factor interactions have significances (coefficient magnitudes) that are similar to several of the two-factor interactions. They are even similar in magnitude to the main effects C and D.

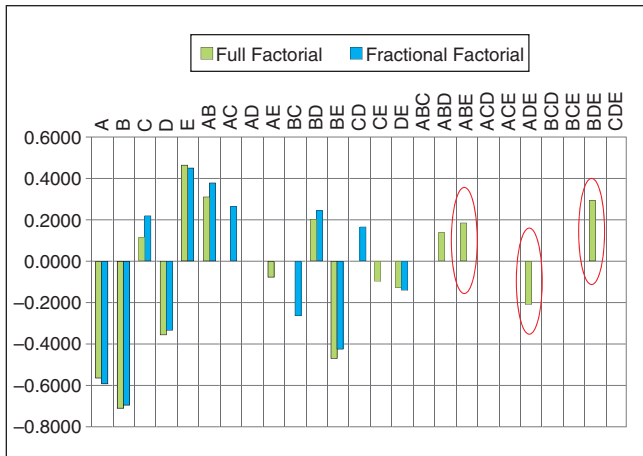
Table 4. Final summary of the full- and fractional-factorial analyses.\*

Term	Full Factorial Coefficient	Fractional Factorial Coefficient	Fractional Factorial Alias
Constant	2.9128	2.8776	
A	-0.5580	-0.5861	
B	-0.7022	-0.6943	
C	0.1163	0.2176	
D	-0.3539	-0.3319	
E	0.4638	0.4491	
AB	0.3091	0.3773	CDE
AC		<b>0.2472</b>	<b>BDE</b>
AD	-0.0006		BCE
AE	-0.0743		BCD
BC		<b>-0.2592</b>	<b>ADE</b>
BD	0.2052	0.2437	ACE
BE	-0.4633	-0.4217	ACD
CD		<b>0.1688</b>	<b>ABE</b>
CE	-0.0931		ABD
DE	-0.1211	-0.1357	ABC
ABC			
ABD	0.1379		
ABE	0.1849		
ACD			
ACE			
ADE	-0.2049		
BCD			
BCE			
BDE	0.2927		
CDE			
R-square	0.9841	0.9920	
Adjusted R-square	0.9672	0.9729	
SST	55.9367	27.9833	
PRESS	4.04416	4.27763	
<b>R-square(PRESS)</b>	<b>0.92770</b>	<b>0.84714</b>	

\* The bolded three-factor interactions were incorrectly identified as significant two-factor interactions by the half-fraction analysis. R-square(PRESS), or the ability to predict new data points within the experimental region, is higher for the full-factorial model. Note: SST = total sum of squares; PRESS = prediction error sum of squares.

One measure of model robustness is the prediction error sum of squares (PRESS) statistic. The PRESS statistic is a measure of how well the model predicts the response within the experimental region. It is calculation intensive, but is easily done using modern statistics software.

Conceptually, the PRESS statistic is computed by fitting the model without one of the data points. That is, we remove



■ Figure 4. Graphical summary of the full- and fractional-factorial analyses. Highlighted are the significant three-factor interactions that were incorrectly identified as two-factor interactions by the half-fraction analysis.

one data point from the set, and then use the remaining  $n-1$  data points to build the model. Then we compute the error (observed – predicted) of the data point that was left out. We repeat this procedure for each of the other data points. We can then sum and square all the errors. This sum of squares is the PRESS statistic.

The larger the PRESS statistic, the worse the model is at predicting new cases. Conversely, the smaller the press statistic, the better the model is at predicting new cases. We can define  $R\text{-square}(\text{PRESS})$  as  $1 - (\text{PRESS}/\text{SS}_{\text{total}})$ .  $R\text{-square}(\text{PRESS})$  is easier to interpret than PRESS alone, since it is similar to the way we interpret  $R\text{-square}$ . The closer  $R\text{-square}(\text{PRESS})$  is to 1.0, the better the prediction ability, or the more robust our model.

In Table 4, the  $R\text{-square}(\text{PRESS})$  has been calculated for each of our two factorial analyses. For the fractional-factorial analysis, the  $R\text{-square}(\text{PRESS})$  is substantially lower than the  $R\text{-square}(\text{PRESS})$  for the full-factorial analysis (0.847 vs. 0.928). This indicates that the model based on the fractional-factorial analysis is less robust, or has a poorer predictive ability, than the model based on the full factorial analysis.

## Conclusions

We have directly compared the results of a full-factorial analysis to what *would have been* the results of a fractional-factorial analysis. This was done for a process that may be considered typical of microelectronics manufacture. In doing so, we have uncovered several significant third-order interactions that are similar in magnitude to several of the second-order interactions — and to a few of the main effects as well.

These results seem to conflict with the popular notion that interactions above second order seldom exist in industrial processes. Again, it should be emphasized that this particular process is not unusually complex relative to many modern processes. In fact, it may be regarded as relatively simple.

Experimenters are encouraged, especially when studying five or more factors, to use a fractional design. Then, if two-factor interactions are observed, the alternate fractions of the experiment can be performed to resolve the higher-order interactions. This is the concept of sequential experimentation. Unfortunately, this second step is rarely performed even when appropriate, because of the sparsity of effects argument. This is particularly true if the model based on the fractional-factorial experiment has a high  $R\text{-square}$  value — as was the case for our process.

Why gather more data if you already have a “good” model? We can draw an analogy to a medical diagnosis: Suppose a doctor spends time and money diagnosing the ailment of a patient. Let’s say the diagnosis is not entirely correct, but the symptoms and lab measurements are consistent with the diagnosis. In addition, the prescribed medicine seems effective.

So in the end all is well — the doctor and patient are satisfied. Would you worry about the patient? Is it possible that as a result of the incomplete understanding, various factors might interact unexpectedly in the future and jeopardize the patient’s well-being? Is it possible the patient’s health could decline in a subtle manner without being noticed?

Again, the experimental design should fit the process, not vice versa. The design should be consistent with the ease of experimentation, the complexity of the process being investigated, and the available resources. The uniqueness of the technology, and the uniqueness of the particular process or phenomenon being investigated should always be considered. In other words, for today’s complex technologies, a one-size-fits-all mentality should be avoided. The decision to perform a screening factorial should be driven by sound business and engineering considerations, not by popular mindset.

In this day and age, the sparsity of effects principle is still good advice. But as educators and practitioners of experimental design, we need to be careful of emphasizing it too much — to the point that it becomes an assumed practice rather than a useful rule of thumb. Otherwise, in our noble effort to simplify and propagate experimental techniques, we gradually lose sight of how to properly and practically tailor each experimental design to the unique process being investigated. **CEP**

**PHILLIP H. WILLIAMS** is a principal staff engineer for Freescale Semiconductor (1300 N. Alma School Road, MD: CH200, Chandler, AZ 85224; E-mail: phil.williams@freescale.com) with responsibilities in quality, consulting, and formal instruction in the areas of statistics and Six Sigma techniques. His 20 years of experience include process and quality engineering in the microelectronics and flat panel industries, where he has held positions in manufacturing, R&D, process integration and factory start-ups. Williams is also a certified Six Sigma Black Belt and Motorola University Master Instructor. He holds a BS and MS in chemical engineering both from Arizona State University. He has been a member of AIChE for the past 22 years.

## Acknowledgements

The author would like to thank the *Freescale Semiconductor* Plating Team for obtaining the experimental data that was analyzed in this article.