Taguchi's Parameter Design: A Panel Discussion

Edited by

Vijayan N. Nair

AT&T Bell Laboratories
Murray Hill, NJ 07974

Panel Discussants

Bovas Abraham and Jock MacKay
University of Waterloo

George Box
University of Wisconsin, Madison

Raghu N. Kacker
National Institute of Standards and Technology

Thomas J. Lorenzen
General Motors Research Laboratories

James M. Lucas
*Du Pont Quality Management and Technology
 Center*

Raymond H. Myers and G. Geoffrey Vining
*Virginia Polytechnic Institute and State University
 and University of Florida*

John A. Nelder
Imperial College, London

Madhav S. Phadke
Phadke Associates, Inc.

Jerome Sacks and William J. Welch
*National Institute of Statistical Sciences and
 University of Waterloo*

Anne C. Shoemaker and Kwok L. Tsui
*AT&T Bell Laboratories and Georgia Institute of
 Technology*

Shin Taguchi
American Supplier Institute, Inc.

C. F. Jeff Wu
University of Waterloo

It is more than a decade since Genichi Taguchi's ideas on quality improvement were intro-
duced in the United States. His parameter-design approach for reducing variation in products
and processes has generated a great deal of interest among both quality practitioners and
statisticians. The statistical techniques used by Taguchi to implement parameter design have
been the subject of much debate, however, and there has been considerable research aimed
at integrating the parameter-design principles with well-established statistical techniques. On
the other hand, Taguchi and his colleagues feel that these research efforts by statisticians
are misguided and reflect a lack of understanding of the engineering principles underlying
Taguchi's methodology. This panel discussion provides a forum for a *technical* discussion of
these diverse views. A group of practitioners and researchers discuss the role of parameter
design and Taguchi's methodology for implementing it. The topics covered include the im-
portance of variation reduction, the use of noise factors, the role of interactions, selection
of quality characteristics, signal-to-noise ratios, experimental strategy, dynamic systems, and
applications. The discussion also provides an up-to-date overview of recent research on
alternative methods of design and analysis.

KEY WORDS: Design of experiments; Dispersion of effects; Location effects; Robust de-
sign; SN ratios; Variation reduction.

Table of Contents

Editor's Introduction

Parameter design (also known as robust design) is a quality-improvement technique proposed by the Japanese quality consultant Genichi Taguchi. It is intended as a cost-effective approach for reducing variation in products and processes. Although it was not introduced in the United States until 1980, Taguchi has been working on this and other quality-improvement ideas for many years dating back to the 1950s. Some Japanese companies and quality-control associations have been using his techniques extensively, although even within Japan his ideas are not universally known or accepted. And before 1980, they were virtually unknown outside of Japan. In 1980, Taguchi received a grant from Aoyama-gakuin University to visit the United States and give lectures on his quality-improvement ideas. He visited several companies and institutions, including AT&T and Xerox. Initial reactions during these and other early visits were generally skeptical, but he managed to capture the interest of a few people. The interest grew, due perhaps to the widespread enthusiasm for Japanese quality practices in the early 1980s. A few individuals at AT&T, Ford, ITT, Xerox, and other places and organizations such as the American Supplier Institute were instrumental in promoting the application of Taguchi's ideas in industry. The first two Mohonk Conferences in 1984 and 1985, organized by the Quality Assurance Center of AT&T Bell Labs, played a big role in exposing his ideas to the statistical community and in stimulating some of the subsequent research. The last 10 years have witnessed much discussion of Taguchi's parameter-design ideas and many applications in industry.

Now that we have accumulated a considerable amount of experience, it seems appropriate to pro-vide readers with a balanced review of parameter design and of techniques for implementing it. There have been many papers and several books explaining, reviewing, or criticizing Taguchi's ideas. Most of these, however, have not adequately captured the diverse views on the topic and their underlying rationale. In particular, the view of the so-called Taguchi school have not been well represented in statistical journals. These considerations led me to organize a "panel discussion" by a group of leading practitioners and researchers. The goal is to provide readers with a balanced and up-to-date overview of (a) the importance and usefulness of the principles underlying parameter design, (b) Taguchi's methodology for implementing them, and (c) the various research efforts aimed at developing alternative methods.

This is different from the usual sort of panel discussion. First, it was not feasible to assemble all participants in a common location. Second, the proceedings of a "free-for-all" oral discussion may have added to existing confusion rather than shed new light on the issues. For these reasons, I solicited comments from participants on a number of topics and created a panel discussion from their comments. Panelists provided comments on topics on which they have worked or with which they have had practical experience. Their comments were organized into sections to give readers a balanced picture of the different views on each topic. Panelists had some opportunity to read and respond to the comments of others. All comments were edited extensively to make the overall discussion and the individual sections flow smoothly and to remove tangential material and excessive overlap. Despite these efforts,

there are still places where the discussion could be smoother or where there is some overlap, but this was unavoidable.

It is not my goal in organizing this discussion to try to resolve any of the differences that currently exist on the issues. The panel discussion is merely intended as a forum for a technical discussion of the diverse views so that readers have a better basis for reaching their own conclusions. Readers should also find the up-to-date overview of recent research efforts and the extensive bibliography useful in obtaining a further understanding of the issues. The discussion has been deliberately kept at a conceptual level and, for the most part, readers have been referred to other sources for relevant technical details. In particular, material already available in *Technometrics* is discussed here only briefly.

To facilitate reading, I have provided in this section a brief summary of Taguchi's parameter-design approach. Readers completely unfamiliar with Taguchi's parameter-design approach should, however, browse through one of the books (for example, Taguchi 1986 or Phadke 1989) or one of the review papers (for example, Kacker 1985 or Nair and Shoemaker 1990) before reading the panel discussion.

Taguchi's philosophy on quality improvement places a great detail of emphasis on variation reduction. Parameter design is intended as a cost-effective approach for reducing variation in products and processes. It can be used either to build quality into new products/processes or to improve the quality of existing ones. Let the product or process under study be referred to as a system. Taguchi classifies the inputs to the system into (a) "control parameters" or "control factors" x—parameters/factors that can be easily controlled and manipulated—and (b) "noise variables" or "noise factors" z—variables/factors that are difficult or expensive to control. Variation in z during manufacturing or operation causes variation in the system's performance measured by some quality characteristic y. There could be many settings of x at which the system can perform, on the average, at desired (target) levels. Among these, there will be some settings at which the system is insensitive to variation in the noise variables z. The basic idea in parameter design is to identify, through exploiting interactions between control parameters and noise variables, appropriate settings of control parameters

at which the system's performance is robust to uncontrollable variation in z. For this reason, the approach is called *parameter design*. The term *design* here refers to the design of a system rather than statistical experimental design. Since the goal is to make the system *robust* to variation in noise variables, the approach has also been called *robust design*.

Taguchi has also proposed a collection of techniques to identify the settings of x that would achieve robust performance. These include statistical experimental design and analysis techniques. The control parameters x are varied according to an orthogonal array ("control" or "inner" array). At each setting of the control parameters, the effects of the noise variables are evaluated by varying them systematically using a "noise" or "outer" array. Taguchi also classifies parameter-design problems into different categories and defines a performance measure, which he calls "signal-to-noise" (SN) ratio, for each category. For example, when the system has a fixed value as the ideal target, Taguchi uses the SN ratio $10 \log_{10} E \, y^2/\text{var} \, y$ as the appropriate measure of variability (see Taguchi and Phadke 1984). At each design setting, data from "replications" across the noise array are used to estimate this measure. The estimated SN ratios are analyzed using standard analysis of variance (ANOVA) techniques to identify the settings of the control parameters that will yield robust performance. Control parameters that do not affect the SN ratio are then used to adjust the average performance on target. Such parameters are called *adjustment factors*, and they may be known a priori or identified through data analysis. The SN ratios and details of the design and analysis vary for other parameter-design problems, but the rationale is similar.

As a check for the assumptions that are implicit in his approach, Taguchi recommends conducting one or more runs at the predicted setting ("confirmation experiments") to verify that the predicted performance is in fact realized.

In addition to the preceding, Taguchi has also proposed a number of techniques for planning experiments (e.g., "linear graphs") and analyzing data (e.g., "accumulation analysis" for ordered categorical data and "minute analysis" for censored data). The panel discussion includes comments on these techniques as well.

1. General Comments

This section contains introductory comments, including the panelists' views on the goals of parameter design and its novelty and importance, and also summaries of their contributions in later sections.

Madhav Phadke

Taguchi's parameter-design method, also known as robust design, is an engineering methodology for improving productivity during research and devel-

opment (R&D) so that high-quality products can be produced quickly and at low cost.

When purchasing a product, a customer considers the features or functions promised by the manufacturer as well as the price. The customer then expects the product to deliver the target performance under all operating conditions, throughout its intended life, without causing harmful side effects. The deviation of the product's performance from the target leads to quality loss and customer dissatisfaction. Robust design is concerned with how to reduce the variation of a product's performance. In particular, it is concerned with selecting the values of control factors (design/process parameters) that minimize the effects of noise factors (uncontrollable parameters). It uses many ideas from statistical experimental design and ANOVA to obtain dependable information about variables in making engineering decisions.

Robust design adds a new dimension to statistical experimental data by explicitly addressing the following concerns faced by all product and process designers:

- How to *economically* reduce the variation of a product's function in the customer's environment.
- How to ensure that decisions found to be optimum during laboratory experiments will prove to be so in manufacturing and in customer environments.

In addressing these concerns, robust design uses the mathematical formalism of statistical experimental design, but the thought process behind the mathematics is different in many ways. In subsequent sections, I will comment on these differences as they relate to the following issues: (1) Role of interactions, (2) selection of quality characteristics, and (3) use of SN ratios for measuring sensitivity to noise factors.

Shin Taguchi

The objective of parameter design is to achieve robust function of the engineering system, either a product or a process, at the lowest costs. Here robustness means that the system performs its function as it is supposed to regardless of various causes of variation. These causes are called *noise*. For example, noises for a paper feeder include paper type, paper size, paper warp, paper surface, paper alignment, stack height, roller wear, and humidity. Robustness to variations in noise is achieved by suitably choosing the settings of the control factors—those factors that the engineer can specify and control with a minimal impact on cost. Fortunately, there are many controllable factors in practice. In the case of a paper feeder, for example, they include roller material, roller

diameter, type of spring, roller contact point, and roller tread design. The objective of parameter design in this case is to find the best combination of controllable factors such that the paper feeder feeds paper properly at a consistent rate under various noise conditions.

Notice that the objective of parameter design is very different from a pure scientific study. The goal in parameter design is not to characterize the system but to achieve robust function. Pure science strives to discover the causal relationships and to understand the mechanics of how things happen. Engineering, however, strives to achieve the result needed to satisfy the customer. Moreover, cost and time are very important issues for engineers. Science is to explain nature while engineering is to utilize nature.

Anne Shoemaker and Kwok Tsui

It should be emphasized that robust design is a problem in product design and manufacturing-process design and that it does not imply any specific solution methods. The goal is to design a system to accommodate a wide range of variation in its inputs. Taguchi deserves credit for pointing out the importance of this problem for producing competitive products.

The solution method appropriate for doing robust design depends on the application area. For example, in mechanical design at General Motors, Hsieh, Oh, and Oh (1990) derived models from laws of physics or geometry. They then developed variance models by Taylor series approximations and used standard optimization methods to find design-parameter levels that minimize variance. See also Box and Fung (1986). In integrated circuit (IC) fabrication at AT&T, underlying physical models are unknown and extremely complicated. Here, fractional factorial experiments are conducted in the laboratory to make processes more robust to factory floor conditions (see Kacker and Shoemaker 1986; Phadke, Kacker, Speeney, and Grieco 1983). In analog-circuit design, on the other hand, circuit simulators are available, so computer experimentation is used (see Buck, Liu, Nazaret, Sacks, and Welch 1989; Nazaret and Liu 1990). Although most of our discussion in later sections will deal with applications in which statistical experiments are used, it is important to keep these distinctions in mind.

George Box

Like most good ideas, designing for robustness has a considerable history. Thus in the early part of this century Gosset, whose product was the barley to be used by the Guinness brewery, emphasized that experiments had to be run in *different* areas of Ireland so as to find varieties and conditions that were in-

sensitive to particular local environments (Gosset 1986). Later Fisher spoke of the "wider inductive basis" for conclusions obtained by comparing treatments within blocks of land that were as different as possible rather than similar. Moreover, the food industry over many years has conducted "inner and outer array" experiments to obtain products such as boxed cake mixes that are insensitive to deviations by the user from the instructions on the box. Youden (1961 a,b) and Wernimont (1977) described methods using fractional factorials for designing analytical procedures that have the property of "ruggedness" so that they would give similar results when conducted in different places and by different people. Even more relevant are papers by Michaels (1964) and Morrison (1957) (see my comments in Secs. 7.1 and 7.2). Although we can still learn from these pioneers, reference to them in no way detracts from the importance of the work of Taguchi in showing the vitally important role that robustness studies can play in the design of industrial products and processes.

The concept that Taguchi calls *parameter design* has many aspects. In subsequent sections, I will address the following: (1) Robustness to environmental variables (Secs. 6 and 7.1), (2) robustness of an assembly to transmitted variation (Sec. 7.2), (3) data analysis for achieving smallest dispersion about a desired target level (Sec. 4), and (4) experimental strategy (Secs. 5.1 and 5.2).

It seems that Taguchi's experimental strategy is intended only to pick the "optimum" factor combination from a one-shot experiment. Although the immediate objective may be this, the ultimate goal must surely be to better understand the engineering system. For example, appropriate designs can provide estimates of those *specific* interactions between environmental and design factors that cause lack of robustness. Once the engineer knows which these are and what they can do, he can employ his *engineering* know-how to suggest ways of compensating for them, eliminating them, or reducing them. Thus I profoundly disagree with the implications of Shin Taguchi's claim that the engineer does not need to "discover the causal relationships and to understand the mechanics of how things happen." To believe this is to discount the way the engineer thinks and the whole basis of his education and experience. It would be a serious mistake to take for granted that such ideas represent the wider Japanese view (for example, see Kusaba 1988).

Thomas Lorenzen

It has been claimed by some that Taguchi invented robust design. At a recent conference in Waterloo

M. F. Franklin told me (parenthetical remarks added by me) that, since the 1940s, work has been ongoing to develop agricultural products that grow uniformly to assure maximum yield (robustness) despite different weather and soil conditions (across noise variables). Franklin said that they like to focus on and use plots of product by weather and soil conditions (control × noise interactions). Sounds like robust design to me! Although the claim of invention is in doubt, there is certainly no doubt that Taguchi has popularized the idea of robustness within the engineering community, and this is a big contribution.

Raghu Kacker

Taguchi's contributions on parameter design can be divided into four categories ranked in order of their merit—quality philosophy, engineering methodology, experimental design, and data analysis. My discussion in later sections will deal with the last three topics. Although the parameter-design approach to variation reduction is clearly very important, it should be kept in mind that it is not a universal approach. In Section 2, I will discuss alternative approaches based on compensation, elimination, and control of the sources of variation.

One of the most significant outgrowths of Taguchi's work is a generalized framework for experiments. He has expanded the traditional scope of designed experiments to cover a wide spectrum of engineering problems. The concepts of both the response variable and the explanatory factors have also been expanded.

The concept of a response variable has been expanded to include various measures of variability (such as SN ratios and other performance criteria), lifetime distributions, and the functional relationship between certain input and output variables. The object of study in parameter design is a measure of variability. In reliability-improvement experiments, the object is the life distribution under various (perhaps stressed) operating conditions. In dynamic problems, designed experiments are used to optimize the functional relationship between an input and an output variable (see Yano [1991] for numerous industrial applications).

Whereas classical statistical designs treat most experimental variables as explanatory factors or block factors, Taguchi (1987, p. 147) has recognized the diversity of the roles of experimental factors. He has classified them, from an engineering viewpoint, into control, noise, signal, adjustment, indicative, and block factors. Each type of factor has an important engineering significance.

This generalized framework for experiments has been instrumental in closing the gap between engineering and statistics.

Jeff Wu

Taguchi's work on variation reduction is widely acknowledged as his most significant contribution to statistics and engineering. The idea is original and its impact profound. He also has some novel ideas in his approach to handling interactions, selection of quality characteristics, and experimental planning techniques. I will discuss the strengths and weaknesses of these ideas in later sections. Many of his techniques for experimental planning were developed in the period 1956–1965, before he invented the idea of parameter design. Some of them were quite original for the time, especially given that he was not in contact with Western statistical literature. Two important papers are those by Taguchi (1959, 1960), which contain several interesting techniques. Most of the details can be found in Volume I of Taguchi (1987).

John Nelder

Taguchi's work on parameter design may be divided into three components—engineering practice, experimental design, and statistical analysis. I leave the assessment of the first to the engineers, who will surely agree that the effect of Taguchi's ideas has been profound. His ideas on experimental design, insofar as they often lead to fractional factorials, are not new, though what is new is the idea of designing for the simultaneous modeling of both mean and variability. In this area, statisticians should be asking themselves why the idea of fractional factorials, known since the mid-1940s, has been so poorly propagated by them. It is in the third component, that of analysis, that statisticians will see room for real improvement. I discuss the statistical defects in Taguchi's analytical procedures in Section 4 and outline in Section 7.4 how generalized linear models (GLM's) provide a general framework for the joint modeling of mean and dispersion.

Jerome Sacks and William Welch

We have few quarrels with Taguchi's parameter-design objectives. There are several features of his formulation and implementation, however, that we do not like. The cross of "inner" and "outer" arrays often leads to a prohibitive number of observations. Moreover, the data from this considerable experimental effort are used very inefficiently. For example, the collapsing of the data to SN ratios surely throws away useful information. Moreover, few of Taguchi's examples consider more than one quality characteristic; this is very unrealistic in our experience. We will elaborate on these issues (Sec. 3), discuss alternative methods (Sec. 6), and illustrate their applications to computer experiments (Sec. 7.3).

James Lucas

The elegance of Taguchi's contributions lies in their essential simplicity. Taguchi has provided a philosophical framework that gets statistically designed experiments run. He uses a loss function to motivate the ideas: Keep a process on-aim, and reduce process variability. This very important contribution is often underemphasized in discussions of Taguchi's contributions (Lucas 1985).

The designs that Taguchi recommends have the two most important characteristics of experimental designs: (1) They have factorial structure, and (2) they get run.

Most of the orthogonal arrays that he recommends are classical screening designs due to Plackett and Burman (1946). Taguchi and his followers have gotten screening designs used much more widely than they were previously. Du Pont's statistical consultants were among the few proponents of screening designs before Taguchi. We know their power and utility. Before Taguchi, however, these designs were seldom used outside the chemical and process industries. Publicizing them, demonstrating their practical power, and getting them used more widely are major contributions of Taguchi.

There is an old consultant's rule that "getting the right design run" gives 90% of the solution. Doing the completely proper analysis is much less important. In many instances, Taguchi has not proposed quite the proper analysis. This has generated some controversy and many papers. By the 90–10 rule, this is a minor criticism.

Raymond Myers and Geoffrey Vining

The use of statistical methods and designed experiments in product and process improvement continues to gain momentum in the United States. Some argue that this is in large part due to, while others assert it is in spite of, the contributions of Taguchi. No one, however, can deny the importance of Taguchi's principles of parameter design. One should design products that are robust to environmental conditions, are robust to component variation, and have minimum variation around a target value. Important questions center on the adoption of these principles by practitioners and the influence that professional statisticians are having on this important subject.

It is probably unfortunate that the important concepts advocated by Taguchi have been overshadowed by controversy associated with his approach to modeling and data analysis. Some of Taguchi's critics have also pointed out the positive aspects of parameter design in the field of quality engineering. An

enormous number of practitioners, however, are still not aware of what Taguchi is all about, though much interest and curiosity persist. Even those who take the time to learn this methodology view Taguchi only through SN ratios, overly simplistic modeling, "pick the winner" analysis, linear graphs, and so forth. For many, the concept of robust products and processes has fallen on deaf ears.

It is clear that the impact of parameter design will be best evaluated after a long period of time. But it is our opinion that the result will depend significantly on how well statistical researchers are able to develop and communicate technology that merges the positive aspects of parameter design with conventional methodology. Only then will the confusion among users slowly cease.

2. Variation Reduction Through Parameter Design and the Role of Noise Factors

Section 2.1 deals with the importance of variation reduction and the role of parameter design, as well as other methods for achieving it. Section 2.2 discusses Taguchi's treatment of noise factors and some of the important considerations underlying their use in variation reduction experiments.

2.1 VARIATION REDUCTION

Raymond Myers and Geoffrey Vining

Before Taguchi's introduction of parameter design in the United States in the early 1980s, our communication of statistical methods to the engineer did not deal sufficiently with the role of product or process variance. There was (and unfortunately still is) a single-minded concentration on the mean of the response of interest. Parameter design can be used to communicate to users that they must consider all sources of process variability. The appropriate set of operating conditions must minimize variability while bringing the mean to target or provide the proper balance between the mean and the variability. The classical assumption of homogeneous variability should only be made when it is truly convincing.

Taguchi's use of SN ratios to capture variability has been a subject of controversy (see the discussion in Secs. 3 and 4). There have been efforts at understanding his ideas and developing more statistically efficient alternatives (Box 1988; Leon, Shoemaker, and Kacker 1987; Nair and Pregibon 1986). Alternative methods proposed include the use of data transformations (Box 1988; Nair and Pregibon 1986; see also Sec. 4), and the use of GLM's for the joint modeling of mean and dispersion (Nelder and Lee 1991; see also Sec. 7.4).

Taguchi's emphasis on variability has also sparked research in the area of dispersion effects and variance modeling. No one can or should imply that Taguchi introduced variance modeling anymore than he invented the notion of squared error loss. The atten-

tion drawn to these concepts by the Taguchi approach, however, certainly influenced Box and Meyer (1986), Nair and Pregibon (1988), Carroll and Ruppert (1988), and many others. It is interesting that after the now classical paper by Bartlett and Kendall (1946), very little appeared that dealt with modeling and controlling process variance until Taguchi. Incidentally, Beckhofer (1960) made an often overlooked contribution in this area. It is important that this way of thinking continues to be reflected in courses taught in the university, as well as in an industrial setting. Academicians are revising courses in linear models and experimental design to more completely accommodate the importance of handling variance heterogeneity.

Raghu Kacker

There are many approaches to reducing performance variation. Taguchi (1987) argued that parameter design is the preferable approach because it involves changing the nominal values of product parameters, which is often cost neutral. Tolerance design is less desirable because it involves using better grade materials and tighter tolerances, which increases the cost. The effective and efficient approach, however, would depend on the nature of the sources of variation (noise factors).

Almost all improvements require two distinct steps—diagnosis and remedy. (Juran [1979] calls this a universal sequence of improvement.) In case of variation reduction, diagnosis means identifying the sources of variation and remedy means instituting countermeasures. Often the most effective countermeasure against variation caused by people, machines, and methods is compensation, elimination, and control of these sources of variation.

Compensation for unavoidable noise is a well-established engineering approach. For example, in automobile manufacturing a chronic problem is the

imbalance of the wheel and the tire. So a U.S. automobile company identifies the heavy point of the wheel and the light point of the tire and then matches them before mounting the tire on the wheel. This reduces the need for final balancing and hence the cost.

Identifying and eliminating (or controlling) the sources of variation is also a basic engineering approach. For example, particulate contamination is a serious problem in IC fabrication, and the most effective and economical approach is elimination and control of contaminants. It is not possible to make an IC-fabrication process insensitive to particulate contamination by parameter design. Bowman, Hopp, Kacker, and Lundegard (1991) described another example relating to the assembly of video cassette recorders (VCR's). Matsushita Corporation has reduced the defect rates in VCR assembly to parts per billion levels by elimination of virtually all sources of variation.

Parameter design is not a universal approach. It relates to those special causes of variation whose effects can be mitigated by changes in the control-factor settings. The success of parameter design depends on two conditions. First, certain interactions between control and noise factors must exist. Second, the engineer is able to identify those factors that are involved in such interactions. Such opportunities may not be rare, but certainly they are not universal.

2.2 THE ROLE OF NOISE FACTORS

Shin Taguchi

The way noise factors are treated is a key concept in parameter design. They are systematically introduced using designed experiments so that their relationships with control factors can be studied. The interactions between control factors and noise are used to reduce variability. Even weak interactions contribute greatly to variability reduction.

When there are many noise factors, it may be difficult to study all of the effects. In this case, we can use compound noise factors that measure extreme conditions of the noise variables. For the paper-feeder example I discussed in Section 1, one can define a compound noise factor with two settings—$N1$ = light/ rough paper and high stack height with new roller and $N2$ = heavy/smooth paper and low stack height with worn roller. In general, the two settings of the compound noise factor, $N1$ and $N2$, are prepared such that they capture variations in other noise factors. The most robust design for the compound noise factor tends to be also robust to all noises.

Jeff Wu

Taguchi's parameter design for variation reduction is a very novel approach. He has advocated the use of a noise (outer) array to systematically vary the noise factors; the noise array is crossed with the control (inner) array, and the product array is used for experimentation. Let m and n denote the size of the control and noise array. Then the product array has nm runs, which can be very large. Recognizing this problem with the product-array format, Taguchi later suggested using a compound noise factor to reduce the size and cost of experiment. It chooses 2–3 (rather than n) extreme level combinations for the noise factors. For this to be effective, however, it requires some rather restrictive and often hard-to-verify conditions, such as (a) the noise effects on the response are unidirectional and (b) the unidirectionality is independent of the settings of the control factors. For details on these conditions, see Phadke (1989).

As a more economical alternative, several authors, including Shoemaker, Tsui, and Wu (1991), have suggested the use of a combined array format that uses a single array to accommodate both types of factors (see the discussion in Sec. 6). The run size of the combined array can be much smaller than a product array. Since run size is not a good measure of experimental cost, I use the cost of runs as a more realistic criterion for comparison. Denote the cost of a control run and of a noise run by c_1 and c_2, respectively. If c_1 is much larger than c_2, the product array format is quite economical. Otherwise, the combined array format is preferred. It is not uncommon that the noise runs are much cheaper. This is why many such experiments have been successful.

Raymond Myers and Geoffrey Vining

Taguchi's use of noise variables is a vital contribution. If important noise variables are used in the experimental process, the variability reflected is that which is most realistic—namely, that which the product experiences as it exits the production line or is used in the field. Parameter design is used to exert control over this type of variability. A chemical engineer cannot report an optimal product blend without accounting for variability produced by different solvents used by various customers. A tobacco chemist must account for variability in quality due to varying storage schedules that may be unforeseen. Any process must address undesirable variability resulting from an inability to control production factors as well as environmental factors. It is true that noise variables did not begin with Taguchi. They enjoyed limited use in this country in foods, gasoline blending, the aircraft industry, and others. But Taguchi dem-

onstrated the advantage of the formal purposefulness of their use as a part of the experimental design.

We often rely on randomization to capture the natural variability of the process. An engineer often intuitively knows that variability is not constant across the design levels. If noise variables are important but not a systematic part of the experimental design, a major portion of variability may be captured haphazardly, indirectly, and thus inefficiently. Taguchi's outer array is meant to characterize process variability at each point of the inner array. It should be pointed out that the use of the inner and outer array does not negate the notion of randomization. In fact, it renders randomization even more important. There is likely to be slippage in complete randomization when noise variables are used. As a result, there is a danger of treating, for the sake of convenience, the conditions in the inner array as blocks or rather whole-plot effects in a split-plot type of design. If this is necessary, then an alternative analysis should be used, as Box and Jones (1992) pointed out, and should be communicated properly to practitioners. Unfortunately, notions of split-plot designs are not a standard part of the engineer's toolbox.

Anne Shoemaker and Kwok Tsui

Since reducing variability is the objective of robust-design experiments, a high priority needs to be placed on careful introduction of noise factors in the experiment. It is essential to identify major noise factors before conducting the experiment, perhaps using a process capability study (see AT&T 1956). These major noise factors should then be systematically varied in the experiment. This active introduction of noise is more efficient than replication and allows us to separately examine variability contributed by each noise factor.

Noise factors that cannot be directly varied in the experiment can sometimes be indirectly varied through surrogates. For example, temperature and gas composition variations inside a reactor can be studied using position in the reactor as a surrogate noise factor (see Kacker and Shoemaker 1986).

Although many remaining small noise effects would be reflected in replication error, replicating every run is frequently not an efficient use of the experimental budget. A more efficient way to gauge repeatability of experimental results might be to replicate only one or two runs.

Because noise-factor levels do not represent a random sample from the noise-factor distributions, sample variances calculated over a noise array are not good estimates of population variances. Instead, they should be interpreted as rough measures of sensitivity to the noise factors present in the noise array.

Bovas Abraham and Jock MacKay

As noted by some of the previous discussants, there are important considerations underlying the use of noise factors in variation-reduction experiments. These will be the subject of our discussion. Our comments are set in the context of *improving* the quality of *existing* products or processes. Although many of these comments are also applicable to the design of *new* products and processes, it is important to keep in mind that there are several important differences between the two contexts.

Meaning of Variation. There is considerable confusion over the meaning of variation. This is apparent in examples described by Taguchi as "the larger the better" and "the smaller the better."

1. One aspect of variation is deviation from target. For example, if a shaft is turned, a measure of quality is the run-out (a measure of out-of-roundness). The ideal value is 0. A single measurement on one shaft may show a deviation or variation from this target value.

2. Part-to-part consistency is another aspect of variation. Large part-to-part variation in run-out may cause difficulties in setup of subsequent operations or assembly.

3. A third aspect is within-part variation. For example, it may be important that the run-out measured at each end of a shaft is close to the same.

A single run-out reading may include all three aspects of variation as well as measurement error.

In simple situations, a statistical model for a response y is

$$y = f(\text{process inputs}) + e, \qquad (2.1)$$

where f is deterministic and the random error e has mean 0 and standard deviation that may depend on the process inputs. Confusion may arise because each aspect of variation can contribute to one or both components in the model. This seems particularly true in situations like the shaft run-out. Is the problem of excessive variation captured by the deterministic or random (or both) components of the model?

With appropriate data, measures can be defined to estimate the different aspects of variation. These measures of variation can be combined into a single performance measure or loss function. The causes of the different aspects of variation, however, may be very different, and it is easiest to search for these causes using separate experiments or at least separate analyses. This point was well made by Box (1988) and several of the discussants to his paper.

Cause of Variation—Noise Factors. Process factors that cause variation in the output/response were

called noise factors by Taguchi (1986). The definition is clear if a factor varies from time to time and transmits piece to piece variation or within-piece variation (because the factor affects different parts of the piece differently). In statistical process control language, this factor is a special cause of variation.

Another type of noise factor corresponds to what Shewhart (1931) called a Type II special cause. These factors do not vary and yet are responsible for some aspect of variation. For example, if shafts are turned on two spindles and there is a systematic difference in run-out between the spindles, then spindle is a noise factor. A fixed factor that causes a systematic difference within a piece is also a noise factor. In terms of the additive model (2.1), these noise factors enter through the deterministic component.

The key strategy that has been suggested for reducing variation caused by a noise factor is to exploit an interaction between the noise factor and some control factor. This interaction must be of the nature that flattens the relationship between the response and the noise factor; that is, the effect of the noise factor within a new setting of the control factors must be significantly less (in absolute value) than the effect within the current setting. In some situations, there is systematic within-piece variation (e.g., differences across locations), but the variation cannot be attributed to any observable noise variables. In such cases, one may be able to identify a dummy or surrogate noise variable (location in this example) and study the interaction between this noise variable and control factors.

Experiments With Noise Factors. Juran and Gryna (1980) suggested that to resolve any problem we should follow a diagnostic journey:

$$\text{problem} \rightarrow \text{cause} \rightarrow \text{remedy}.$$

In this instance, the problem is one of excessive variation and the cause has been identified as one or more noise factors. The goal of the variation-reduction experiment is to find the remedy to the variation associated with these important noise factors. To minimize the experimental effort, it is important that noise factors that contribute substantially to the variation are included in the study. In the production setting, available data can often be used to identify noise factors using stratification, regression methods, and control charts.

Knowledge of the noise factors and their behavior is an important prerequisite to an efficient experiment. Since it is essential to find control factors that interact with the noise factors, the more that is known about how the noise factor acts, the more likely that the control factors and their levels will be successfully chosen. If the noise factor is transmitting variation and the experiment is to be run with an outer array, then understanding of the range of variation of the noise factors is required to select the levels. Sometimes the noise factor cannot be controlled during the experiment, but knowledge of its behavior can be used in the design. For example, in a foundry the pouring temperature of the iron was identified as a noise factor. It was not possible to control the temperature sufficiently to use it in an outer array. Instead, the factor that the temperature could be measured and that it changed rapidly after lunch break was exploited to design an experiment in which control factor by temperature interactions were examined using a linear model for the observed response, which included temperature by control-factor interactions.

In cases in which the noise factors are not identified or measured and controlled during the experiment, it is common to define a run of the experiment so that there is sufficient time for the noise factor to "act." Knowledge of past behavior of the process is critical in defining an experimental run. The process must be stable from run to run so that each run will experience the same variation. In addition, the run must be long enough so that the variation is practically important. Control factors must be selected with little understanding of why the variation is occurring. Quinlan (1985) gave an example in which shrinkage was measured at four different places within a length of cable. No noise factors were explicitly identified to explain the excessive variation in shrinkage. In this situation, there was little choice but to model performance measures (SN ratios, variances, etc.) calculated for each run.

It is advisable, however, to try to explicitly identify the noise variables and model directly the observed response to examine the control by noise-factor interactions (see Sec. 6). This is preferable to computing SN ratios or variances and modeling dispersion effects. Given the limited amount of data usually available, estimating dispersion effects precisely is likely to be difficult.

3. The Role of Interactions, SN Ratios, and Selection of Quality Characteristics

Taguchi's philosophy on interactions has been the subject of considerable debate. In this section, some of the discussants explain Taguchi's philosophy on interactions and his selection of SN ratios and quality characteristics. Others assess the validity of these views.

Madhav Phadke

The role of interactions has been debated vigorously since Taguchi's approach to experimental design for designing robust products and processes became known in the United States. A perception persists in the statistical literature that Taguchi's approach assumes that interactions are absent and hence the method is unscientific. The lack of adequate literature in the English language, the evolving nature of the methodology, and the lack of understanding of the engineering issues on the part of statisticians have been responsible in part for the misunderstanding and debate. An engineering perspective of this debate is given in the following.

In designing robust products/processes, one must first divide all factors that affect the product's output into two categories, control factors (C) and noise factors (N). The interactions among these factors can be divided into three categories—among control factors $(C \times C)$, between control and noise factors $(C \times N)$, and among noise factors $(N \times N)$. During parameter design, one is interested in choosing the levels of control factors so that the product's response is least sensitive to noise factors and can be adjusted on target as appropriate. The $C \times N$ interactions are exploited to accomplish this. The $N \times N$ interactions play little role in making a product's performance insensitive to noise factors.

What is the role of $C \times C$ interactions in reducing the sensitivity of a product's response to noise factors? Do $C \times C$ interactions exist? If so, how should they be handled? These questions form the source of controversy and debate.

Taguchi's robust-design method addresses the problem of interaction among control factors in a way that is philosophically different from the classical approach to experimental design. Presence of large $C \times C$ interactions is considered highly undesirable for several reasons:

1. First, presence of interactions implies that a much larger number of experiments would be needed to study the same number of control factors.

2. Second, the presence of large $C \times C$ interaction makes it difficult to divide the task of designing a complex product into several smaller tasks (subsystems) that could be investigated simultaneously by different teams of engineers. This is highly undesirable for shortening the development interval and for improving R&D productivity. Moreover, this makes it difficult to reuse the subsystem design for other products. Consequently, overall R&D costs are higher and the development intervals longer.

3. Third, and most important, the reason for seeking additivity has to do with the ability to transfer designs from laboratory to manufacturing and eventually to the field. The conditions under which the experiments are conducted can also be considered as a control factor with three settings—laboratory, manufacturing, and customer usage. If strong $C \times C$ interactions are observed during laboratory experiments, these control factors are also likely to interact with conditions of experimentation. In this case, optimum settings in the laboratory may not prove to be optimum under manufacturing or customer-use conditions. Thus the manufacturing reject or rework rate may turn out to be high, cost design changes may become necessary, the product may fail in the field sooner than expected, and the product may not function on target under different customer environments.

Thus every attempt is made in robust design to eliminate or minimize the $C \times C$ interactions through judicious choice of the quality characteristics (responses used in robust design), the objective functions to be maximized (the SN ratios), the control factors, and their levels (including the use of "sliding levels" of factors—see Phadke [1989, p. 145] for an example). Orthogonal arrays and confirmation experiments are used as a method to check for additivity (see Sec. 5.1 for more discussion of this topic). Choosing these quantities properly often constitutes the bulk of the effort in planning robust-design applications. These tasks require engineering know-how about the specific project and also knowledge of robust-design methodology. Care taken in this activity can greatly enhance the ability of the robust-design experiment to generate dependable and reproducible information with a small number of experiments. If an engineer is unable to eliminate interactions, he must continue to research different problem formulations or accept the risk of sending defective designs to the next product realization stage. There are no rules that can guarantee absence of interactions.

This must be achieved on a case-by-case basis and even then sometimes by trial and error.

Here are some important guidelines for selecting the quality characteristics to minimize interactions:

1. Identify the ideal function or the ideal input–output relationship for the product or the process. The quality characteristic should be directly related to the *energy transfer* associated with the basic mechanism of the product or the process. Avoid focusing on the ways energy is wasted.

2. As far as possible, choose *continuous variables* as quality characteristics.

3. The quality characteristic should be *monotonic*; that is, the effect of each control factor on robustness should be in a consistent direction, even when the setting of control factors is changed. In several situations, it is difficult to judge the monotonicity of a quality characteristic before conducting experiments. In such situations, one has to conduct experiments followed by the confirmation process to determine if the quality characteristics have monotonicity.

Additional guidelines for selecting the quality characteristics and the SN ratios and examples were given by Phadke (1989) and Phadke and Taguchi (1987). These guidelines include both what to do and what not to do. Finding quality characteristics that meet all of these guidelines is sometimes difficult or simply not possible with the technical know-how of the engineers involved. The robust-design experiment, however, will be inefficient to the extent that these guidelines are not satisfied. It is a common mistake to use percent defective or yield as quality characteristic. It violates the rules described previously and should be avoided.

Shin Taguchi

In parameter design, the most important job of the engineer is to select an effective characteristic to measure as data. For example, a coating process results in various problems such as poor appearance, low yield, sags, orange peel, and voids. Too often, people measure these characteristics as data and try to minimize or maximize the response. This is not sound engineering, because these are simply the symptoms of poor function. It is not the function of the coating process to produce an orange peel. The real problem is the functional variability of the coating process due to noise factors such as variability in viscosity, ambient temperature, and coating surface variability. We should measure data that relate to the function itself and not the symptom of variability.

To determine an effective characteristic, it helps to consider the underlying transformation of energy in the engineering system. Quality problems take place due to variability in the energy transformation. Considering the energy transformation helps to recognize the function of the system. One fairly good characteristic to measure for the coating process is coating thickness. After all, the function of the coating process is to create the coating layer. Symptoms such as orange peel and poor appearance result from variability of coating thickness. It is sound engineering strategy to measure the coating thickness and to find the combination of controllable parameter settings such that its variability is minimized.

The efficiency and effectiveness of engineering activities depend greatly on what is measured as data. In general, attribute characteristics and others dealing with notions of "smaller-the-better," or "larger-the-better" are only symptoms of variability/performance. They tend to introduce strong, adverse interactions. It is better to use energy-related nominal-the-best type characteristics. It is even better to use dynamic characteristics (see the discussion in Sec. 7.6).

Thomas Lorenzen

The statement that causes me immeasurable grief in consulting with engineers is: If the response variable is chosen to reflect either energy output or fundamental physics, there will be no control-factor interactions, so none need be considered in the design or analysis of the data. This is particularly grating to me because it is so easy to remember, and everyone wants life to be easy and require less work.

I have several problems with this statement. First, although different response variables will most assuredly influence the complexity of the required model, neither measurable energy output nor fundamental physics need be additive. Sorry about that, but interactions may be needed. Second, control × noise interactions are necessary to improve robustness. The difference between a control and a noise factor is definitional, whether the factor can be controlled inexpensively in the factory or not. I am left with the conclusion that energy output and fundamental physics know whether the factor can be controlled in the factory or not and form interactions appropriately! Finally, one is interested in robustness to the customer, not robustness to energy output or fundamental physics. Before I am comfortable, I need to know the relation between energy output or fundamental physics and the customer-perceived quality characteristic.

The claim that confirmation experiments will tell if the fit is good is also not correct. Recently, an engineer modeled expensive computer runs following a course offered by the American Supplier Institute. The "best point" confirmed. I heard this

presentation and talked him into running a higher resolution design requiring the same number of runs. The "best point" from this model also confirmed, with a 30% improvement! End of claim and an engineer who now believes in interactions.

Raghu Kacker

The justification given for Taguchi's philosophy on interactions is that the design engineer needs to determine, through laboratory experiments, settings of control factors that are optimal in the manufacturing conditions and in the customer's use conditions (Taguchi 1987, pp. 117–141). It is claimed that when the effects of the control factors are additive (i.e., the interaction effects among control factors are negligible in comparison to the main effects), the optimal settings for the laboratory environment are likely to remain optimal during manufacturing and customer-use conditions; otherwise extrapolation to manufacturing and use conditions may not be achieved.

This argument fails to recognize that a design process is rarely done in a vacuum. A well-managed design process should have access to experience with related products and processes. Effective use of such information in the statistical design and analysis of laboratory experiments is the rational approach to extrapolation of results. According to Rosenblatt and Watson (1991), successful companies design for manufacturing and customer-use conditions by employing a concurrent engineering approach that brings to the designers' table accumulated experience from manufacturing as well as accumulated data on the performance of related products in use conditions. I think the use of such prior experience rather than ad hoc philosophy concerning the additivity of control-factor effects is the key to extrapolation of laboratory results.

Taguchi (1987, pp. 59–61) pointed out that the response (or performance) characteristic is not given; it is chosen. He emphasized that the response characteristic should be chosen such that the effects of the control factors are additive. In my view, it is unrealistic to assume that one can always find relevant performance characteristics that are additive in the effects of control factors. I agree that a study of the underlying mechanism may be more effective than the direct focus on final characteristics. But interactions may still be present in the underlying mechanism. Even if additive characteristics exist, say at molecular levels, it may not be possible to identify and measure these characteristics within the scope of the experiment.

It is frequently a matter of scientific research to identify relevant performance characteristics. For example, in materials research engineers try to identify and select characteristics that are likely to have the largest effect on the desired properties of the product. The identification of such characteristics is often hampered by measurement error. When the selected characteristic cannot be measured with reasonable precision, a surrogate characteristic (whose relation to the characteristic of interest is fairly understood) is used.

In addition, sometimes there is no accepted standard for measuring a response characteristic. For example, to measure the hardness and fracture toughness of ceramic composites, a diamond wedge is indented into the composite (for example, see Fuller et al. 1991). The defined characteristics are functions of the depth profile of the indentation and multiple cracks that develop, and there is no universally accepted way of calculating these characteristics from the various types of depth and cracks that can form. An individual engineer can use an ad hoc characteristic. But generally accepted standards for measurement are needed to compare results from different sources.

Jeff Wu

Taguchi (1987, p. 61) stated, "The efficiency of research will drop if it is not possible to find characteristics that reflect the effects of the individual factors regardless of the influence of other factors." In more precise terms, it means that the characteristics should depend only on the marginal effects of the individual (control) factors. Taguchi called such characteristics *monotonic*. To achieve this, Taguchi (1987, p. 171) and Phadke and Taguchi (1987) suggested the following steps:

1. Find a characteristic possessing monotonicity.
2. Use sliding levels when the factors are interrelated.
3. Use an SN ratio as the objective function for analysis.
4. Use a "correct" analysis method—that is, accumulation analysis for ordered categorical data and minute accumulating analysis for censored data.

Generally speaking, 1 is very original, 2 is a useful reminder of what has been known but not emphasized, and 3 and 4 are faulty. Let me elaborate on them.

In the statistical literature, a characteristic y is usually given and unquestioned, and a model is fitted to describe y as a function of some covariates x. A transformation on y (or on both y and x) may be considered to improve the model fit. If the original y does not allow a monotonic relation in x, a data transformation, even if it gives a better fit, will usually not result in a monotonic relationship that holds outside the region covered by x. Therefore, a transformed relationship may not be effective in predic-

tion. A more effective approach is to use the subject-matter knowledge to find a monotonic characteristic. Several interesting examples can be found in chapter 6 of Phadke (1989). One of them is about heat-exchanger design. If τ is the target temperature and $L = |y - \tau|$ is the loss function, it is obvious that L is not monotonic even for monotonic y, since L is not a monotonic function in y. This is why Taguchi advocates that analysis be done on the original scale y instead of on the transformed scale L. Having said this, I have difficulty understanding why he advocates the use of SN ratios for achieving monotonicity. Take the most commonly used SN ratio,

$$\hat{\eta} = \log_{10}(\bar{y}^2/s^2), \qquad (3.1)$$

where $\bar{y}$ and s^2 are the mean and variance of y_j. It is a nonmonotonic and many-to-one transformation of y_j. (Most of his other SN ratios also possess this undesirable property.) There is an apparent contradiction here to what he advocated in Step 1. As argued in 1, it is better to analyze the original response y instead of the SN ratio η (see Sec. 6). Criticisms of his SN ratios abound in the literature and will not be repeated here.

Use of sliding levels to account for the interrelationship between factors and to minimize interactions is a good but often neglected practice. Strictly speaking this is not his original idea. But I think it is fair to say that Taguchi has brought it to our attention and has used it in many of his case studies. (Using case studies to make his points is a very important practice he introduced to the inward-looking statistical community.)

Both accumulation analysis and minute accumulating analysis have been studied and criticized as being unnecessarily complicated and often invalid (Box and Jones 1986, 1990; Hamada, in press; Hamada and Wu 1990; Nair 1986). These studies also

show that they can detect spurious interactions and thus create nonmonotonicity.

Jerome Sacks and William Welch

There seems to be confusion about attitudes to interactions in Taguchi's experiments. Taguchi and Wu (1985, p. 55) said that the engineer should convert the quality characteristic into one having additivity. Ignoring the obvious practical difficulties, suppose that the engineer can carefully parameterize the problem to minimize interaction effects. As a simple illustration, let the quality characteristic, y, depend on two control factors, x_1 and x_2, and a noise factor, z, through $y = x_1 + x_2 + z$. In Taguchi's implementation, however, it is an SN ratio, not y, that is analyzed. When the "replicates" $y_1, \ldots, y_n$, arising from n levels of z, are reduced to, say, the smaller-the-better SN ratio $S = \log (1/n \, \Sigma y_i^2)$, then S is clearly no longer additive in x_1 and x_2. Despite the engineer's efforts, the interaction between x_1 and x_2 cannot necessarily be ignored.

Taguchi's main motivation for ignoring interactions between control factors appears to be economy of experimental effort rather than any assurance that it is safe to do so. Economy measures are forced by the inefficiency of his crossed-array experimental designs. For each combination of control factors in the control (inner) array, he makes observations at all combinations of the noise factors in the noise (outer) array. With m rows in the control array and n rows in the noise array, there is obvious potential for mn, the total number of observations, to become prohibitively large. Therefore, he attempts to keep m (and n) as small as possible by ignoring interactions. The methods we outline later in this discussion, based on a single experimental array for both control and noise factors, typically require far fewer observations *and* allow interaction effects to be modeled.

4. Data Analysis: Use of SN Ratios, Data Transformations, or Generalized Linear Models?

As noted earlier, Taguchi classifies parameter-design problems into different categories and defines different performance measures called SN ratios for each problem. In this section, the panelists consider the important special case in which the product or process has a fixed target value and discuss the use of Taguchi's SN ratio and other, more established statistical methods.

Madhav Phadke

Unlike what has been proposed in some of the statistical literature (see Box 1988; Nair and Pregibon 1986) selecting the SN ratio is not an exercise in determining a data transformation that stabilizes the variance. It is the process of identifying the ideal relationship between the signal factor and the quality

characteristic and evaluating the sensitivity to noise factors with respect to the chosen ideal function. Assuming an appropriate adjustment (such as ability to change the scale or location) is also important because it allows evaluation of sensitivity to noise factors under a set of standard conditions.

I will illustrate the rationale behind the SN ratio for a common type of robust-design problem called the nominal-the-best type of problem. Let τ be the target (nominal value). Then, the quadratic loss associated with the target value τ is given by

$$Q = (\mu - \tau)^2 + \sigma^2, \qquad (4.1)$$

where μ and σ denote the mean and standard deviation of the response variable. Suppose that we know the mean and standard deviation for two different processing conditions. How can we say which processing condition is preferable? How can we evaluate their relative sensitivity to noise factors? For proper comparison, we must first evaluate the quadratic loss after adjusting the means of the two process conditions to the target value τ.

In some cases, it is relatively easy to identify a scaling type of adjustment factor. For example, in the polysilicon deposition process discussed by Phadke (1989) deposition time is an adjustment factor; that is, if the deposition time is changed by a factor r, the mean and standard deviation also change by the same factor. In this case, we can *calculate* what the quadratic loss would be after adjusting the mean μ on target. When the mean is changed from μ to τ—that is, $r = \tau/\mu$—the standard deviation would change from σ to $(\tau/\mu)\sigma$. The corresponding quadratic loss, called the quadratic loss after the adjustment, is given by $Q_a = (\tau/\mu)^2\sigma^2 = \tau^2(\sigma^2/\mu^2)$. Because τ is the same for the different processing conditions, we can compare the sensitivity to noise factors by comparing the corresponding values of (σ^2/μ^2) or its reciprocal. This is equivalent to the SN ratio: $\eta = 10 \log_{10}(\mu^2/\sigma^2)$ [see also (3.1)]. To achieve robustness, the control factors are chosen to minimize this SN ratio. The adjustment factor is then tuned to get the mean on target as needed. This procedure of process optimization is generally called the two-step optimization: (1) Maximize SN ratio, and (2) bring the mean on target. See Phadke and Dehnad (1988), Phadke (1989), Leon et al. (1987), and Leon and Wu (in press) for further discussion of two-step optimization.

When the adjustment factors cannot be identified a priori, experimental data can be used to discover a suitable adjustment factor by estimating the effects of control factors on η and μ. The control factors can then be divided into three categories: (1) Factors that influence η (these factors are useful for improving robustness), (2) factors that influence μ but do not influence η (one or more of these factors can

be used for adjusting the mean on target, (3) factors that do not influence μ or η (these factors can be used to satisfy some other purposes, such as convenience or cost).

It is sometimes tempting to view the problem as direct minimization of the quadratic loss function given by Equation (4.1). When that is done, there is an increased risk of interaction among the control factors as illustrated by Phadke (1989) for the polysilicon-deposition case study. Furthermore, the quadratic loss function Q is dominated by the term $(\mu - \tau)^2$. Hence minimizing Q is not very effective in minimizing sensitivity to noise factors. Indeed, the interactions caused by the term $(\mu - \tau)^2$ can lead to decisions that do not reduce sensitivity to noise factors. When we compute Q_a or η, however, we isolate sensitivity to noise factors. Hence their optimization leads to reduced sensitivity to noise factors.

In robust design, engineering problems are categorized according to the nature of the signal factor and the quality characteristic. Cases in which the signal factor is absent are called static problems, whereas cases in which the quality characteristic must track the signal factor are called dynamic problems (see Sec. 7.6). The preceding principles apply for deriving the SN ratio in each of these cases. Cataloging SN ratios for new types of engineering problems is an important research area. Several commonly encountered SN ratios in both the static and dynamic situations were given by Taguchi (1978) and Phadke (1989).

George Box

Suppose the effect of an adjustment factor on the response y is multiplicative. Its effect on $\log y$ would be additive, however, so one does not have to worry about its effect on the variance of $\log y$ when tuning the factor to get the mean of $\log y$ close to its target value. Thus, despite Madhav Phadke's discussion, it is clear that in this case the log-transformation decouples the dispersion and location effects and so simplifies finding those conditions x_0 that simultaneously locate the process on target and minimize dispersion about the target. In fact, it has been shown that Taguchi's SN ratio for this problem, which is proportional to $\log(Ey^2/\text{var } y)$, is closely approximated by $-\log \text{var}(\log y)$. See Box (1988), Box and Fung (1986), Leon et al. (1987), and Nair and Pregibon (1986).

There is no guarantee, however, that the adjustment factor would always have a multiplicative effect on the response y. A better alternative would, therefore, seem to be to evaluate a range of transformations that might include the log as a special case (and no transformation as another special case) and carry out the analysis in terms of that transformation, which

yields maximum simplification. The method was illustrated for the particular case of the family of power transformations y^λ by Box (1988), who showed how a simple graphical method (called a lambda plot) could be employed to set out the possibilities for simplification.

John Nelder

Taguchi's approach to data analysis begins by defining a summarizing quantity, an SN ratio, and then seeks a model for it in terms of the experimental factors. As has been discussed in the literature, there are often serious objections to the forms of his SN ratios. Their use can also lead to great loss of information (in the statistical sense) in an analysis and so fail to use all of the information in the data.

There is a more general objection to this way of proceeding—namely, that it inverts the processes of model selection and model prediction, where the latter term is used to mean the formation of summarizing quantities, and estimates of their uncertainty, from the parameters estimated during the model-selection process. A typical example of a summarizing quantity is the estimated LD50, or median lethal dose, from the results of a quantal bioassay. Having fitted a linear model on the probit scale with log dose as the explanatory variable and yielding parameter estimates b_0 and b_1, we estimate the LD50 as $-b_0/b_1$. Here it would not be possible to use Taguchi's approach by estimating the LD50 from each response unless the slope of the line was known a priori, but the general principle holds: First fit a model to suitable responses (model selection) and then form the quantities of interest (model prediction) from the parameter estimates. An empirical reason for not forming the summarizing quantity first is that models

for it are often more complicated than those for the basic responses.

Another objection to the analysis of the SN ratio is that it preempts the definition of the summarizing quantity, whereas in real life the definition may well depend on what the analysis shows. Although arguments based on loss functions are certainly relevant here, the experimenters need to consider what is appropriate to the special circumstances of their production system and form their summarizing quantities accordingly. Although it is attractive to some that the use of a standard SN ratio avoids the necessity of thinking about their experiment, avoidance of thought, as usual, does not pay in the long run.

The use of data transformations has been suggested as a better alternative to Taguchi's SN ratios (Box 1988; Nair and Pregibon 1986). This method seeks a transformation of the data, $f(y)$ in place of y, with the aim of fulfilling two criteria, separation and parsimony. Separation means that the transformation should eliminate any unnecessary complication in the model due to functional dependence between variance and mean, and parsimony means that the transformation should provide simple additive models for the mean and dispersion. It is asking rather a lot of a transformation that it should produce simultaneously separation, additivity, and an approximately normal error structure, and indeed there will be many cases in which it cannot, particularly when the data have the form of counts or ratios of counts (proportions). I will show in Section 7.4 that the use of GLM's removes many of these difficulties and in particular that it integrates the analysis of counts and proportions with that of continuous responses. With these models the behavior of the mean and variance can be modeled quite separately. Furthermore, when we use a GLM, we do not transform the data, so its original dimensions are preserved throughout.

5. Experimental Strategy and Planning Techniques

5.1 EXPERIMENTAL STRATEGY

Taguchi's experimental strategy consists of (a) running highly fractional experiments using orthogonal arrays and analyzing the data to identify appropriate control parameter settings and (b) running a confirmation experiment to verify that robust performance is achieved at the identified parameter settings. Some discussants elaborate on the rationale for this approach, and others argue for the use of a sequential experimental strategy.

Madhav Phadke

In this section, I will discuss the strategy used in robust design for selecting the control arrays and how it is related to the overall philosophy of interactions I discussed in Section 3.

Although the quality characteristic and the SN ratio for the experiment should be carefully chosen based on engineering knowledge so as to minimize control-factor interactions, it is still necessary to establish that interactions are small or absent. To do

this, the levels of the control factors are varied according to suitable orthogonal arrays. These arrays are chosen so that interactions are deliberately confounded with the main effects. Using only main effects, the SN ratio is predicted under conditions other than those in the orthogonal array and compared to the results from confirmation experiments. If strong interactions are present, predictions would not match confirmation experiments and we would detect the lack of additivity. A search is then made for a new SN ratio with appropriate new adjustment or a new quality characteristic is investigated. This process is continued until the additivity of SN ratio is established. Of course, the more experience an engineer has with robust design, the fewer iterations he/she would need.

Raghu Kacker

The orthogonal arrays (of strength 2) popularized by Taguchi satisfy a niche in between one-factor-at-a-time plans and experiments for scientific feedback. Industrial experiments have two important roles— geometrically balanced coverage of the experimental region and prediction beyond the actual tests conducted. Statistical modeling is not necessary for experiments to be useful. Sometimes the following steps suffice: (1) Visualize that k control factors form a k-dimensional experimental region. (2) Judiciously select m test points and conduct the experiment. (3) Pick the winner of the m test runs (or foretell winning conditions based on the results of m tests in light of subject matter knowledge. (4) Confirm that the winning conditions are better than the reference conditions.

These steps are fairly close to what many engineers do; they use personal judgment in selecting the test points and make ad hoc conclusions in view of simple plots of the data and their subject-matter knowledge. For doing step (2), orthogonal arrays provide a geometrically balanced coverage of the experimental region. This simple approach makes no assumptions about the complexity of the response surface, and only simple plots of the data are needed. It beats the popular one-factor-at-a-time practice and may be useful as an initial approach leading to more sophisticated strategies.

A National Institute of Standards and Technology (NIST) computer scientist found this pick-the-winner-of-an-orthogonal-array experiment approach useful in a computer experiment. He used an $OA_{25}(5^6)$ to identify test settings that reduce a response (see Lyon, Snelick, and Kacker 1991). Subsequent experiments, statistical modeling, and prediction have not yet beaten the minimum obtained with the initial orthogonal-

array experiment; the response surface appears very nonlinear and complex.

Jeff Wu

Taguchi attaches great importance to confirmation experiments in his design strategy—that is, a small follow-up experiment to confirm the findings from analysis of experimental data. Distinction should be made between traditional response-surface methodology, which commits more runs to regression model building, and this approach, which commits more runs directly to confirmation. The former is preferred if it is important and affordable to understand the response relation and if the relation is not too complex. The latter is preferred if it is important to quickly identify a setting (rather than a whole surface) with a better performance. This latter scenario is commonly encountered in solving practical quality problems. We should point out, however, that the analysis of marginal means, which is Taguchi's main strategy for confirmation, has serious problems. It is only justified when the characteristics are monotonic. Otherwise, it can miss important (synergistic) interactions and lead to poor prediction of optimum settings. For an investigation of its deficiencies and some remedial measures, see Wu, Mao, and Ma (1990). Sound strategies for confirmation experiments are needed. Possible approaches include response-surface methodology (based on model fitting) and search methods such as sequential elimination of levels (see Wu et al. 1990).

George Box

As I stated in Section 1, Taguchi's experimental strategy seems intended only to pick the "optimum" factor combination from a one-shot experiment with the addition of one or more confirmatory experiments (whose value has been overrated; for example, see Bisgaard and Diamond [1991]). The ultimate objective of the experimental investigation must surely be to better understand the engineering system. To do this requires, I believe, efficient statistical tools of design and analysis that accommodate the naturally iterative process of scientific method characterized for example by the Deming–Shewhart cycle.

The beginning of an investigation, when the engineer may be required for example to "list all the important variables and their important ranges of variation," is the time when he/she knows least about the problem. In an iterative procedure, having performed an initial limited number of runs, simple analysis aided by computer graphics can allow him/her to mull over the effects induced simultaneously in the various measurements in relation to basic engineering know-how. This in turn can suggest how the investigation should proceed and lead to a new ex-

perimental design, which, because of what has been learned may include quite different questions, different choices of factors, and even different measurements from those considered at the first stage.

We have a large reservoir of engineers with a vast background of engineering know-how. They need to learn statistical methods that can tap into that knowledge. Statistics used as a catalyst to engineering creativity will, I believe, always result in the fastest and most economical progress.

Raymond Myers and Geoffrey Vining

Taguchi's analytical methodology leading to optimum conditions leaves practitioners with the impression that the sound statistical analysis is a "one-shot" operation. We feel that this ignores the important lessons from Box and Wilson (1951) and classical response-surface methodology about the virtues of planning experiments sequentially. In particular, we are referring to the stages of variable screening, region movement, design augmentation, the fitting of a more elaborate model, and finally the exploration of the experimental region via response-surface methods. One phase of a total experimental plan dictates the succeeding one. It is our feeling that users are learning about methodology with no regard to the proper context of usage. The details of a fractional factorial may be well known by the practitioner, but he or she may not be aware of where in the total study it should be used. For example, if dispersion effects are found using a fractional factorial, what comes next?

There are certainly a number of success stories in industry in which parameter design was used effectively, both with and without the use of analytical methods introduced by Taguchi. This should not be surprising. The total approach includes the use of a factorial structure of experimentation. As a result, even the use of less than the most appropriate methods can certainly produce positive results.

It is our opinion that there should not be a preoccupation with the singular goal of finding *an estimate* of optimum conditions. Too often, the engineer will have sound, pragmatic reasons why a single-point estimate of the optimum cannot be adopted. We feel that the sequential approach should be resurrected with vigor, in harmony with sound modeling and the important principles of Taguchi's parameter design. The area has yet to reach a stage of maturity, with efficient techniques having not filtered down to the user. Moreover, these methods will not be practical for the user until appropriate software is developed and widely distributed.

5.2 EXPERIMENTAL PLANNING TECHNIQUES

Jeff Wu

I will focus here on two of Taguchi's experimental planning techniques, which in my opinion are either original or have important practical applications. A third technique, using idle columns for generating nonorthogonal arrays, was studied by Grove and Davis (1991).

Mixed-Level Orthogonal Arrays With Economic Run Size. Interest in orthogonal arrays has traditionally been focused on the 2^{n-p} and 3^{n-p} fractional factorial designs defined by a subgroup of defining relations. See, for example, the influential texts by Kempthorne (1952) and Box, Hunter, and Hunter (1978). There are at least two problems with them. First, there are gaps between the run sizes of these arrays. Second, these arrays with mixed levels can be very large. With this in mind, let me now ask why Taguchi favors the use of the following arrays: $L_{18}(2 \cdot 3^7)$, $L_{18}(6 \cdot 3^6)$, and $L_{36}(2^{11} \cdot 3^{12})$. The main reason is run-size economy. For five to seven 3-level factors, the best among the 3^{n-p} series has 27 runs. Use of $L_{18}(3^7)$ results in a 50% saving of runs. Justifications for the other arrays are quite obvious. Partly stimulated by the increasing use of these arrays, Wang and Wu (1991) developed a general approach to the construction of mixed-level orthogonal arrays with economic runs. The article contains a good collection of these arrays with less than 100 runs. Several methods have been proposed in the combinatorial-design literature for constructing mixed-level arrays, but the emphasis is not on run-size economy. For any combinatorial work to make impact in industrial applications, this and other practical constraints cannot be overlooked.

How should data from such arrays be analyzed? Unlike the 2^{n-p} and 3^{n-p} designs, whose effects are either orthogonal or fully aliased, these arrays have more complex aliasing patterns (the most notorious being the 12-run Plackett–Burman design.) The traditional approach is to use them for screening only. Hamada and Wu (1992) argued that, because in many practical situations few main effects and fewer interactions are important, it is possible to entertain and estimate the important interactions. They proposed an analysis method for doing this and demonstrated its effectiveness on several real experiments. Taguchi's view on this issue is different. Taguchi and Wu (1985, p. 35) stated, "no interactions are calculated even if they exist. . . . these interactions are treated as errors, so it is advantageous to have the effects of these interactions uniformly distributed in all (design matrix) columns." From this and other statements made elsewhere, Taguchi seems to be-

lieve that estimated main effects are not affected by interactions because they are smeared or evenly spread across all of the design matrix columns. The results of Hamada and Wu (in press) show the contrary (see also Box 1952).

Linear Graphs. In planning an experiment, prior knowledge may suggest that some interactions are potentially important and therefore should be estimated clear of the main effects. This will not be a problem if a resolution V design can be employed. Quite often a smaller design (resolution III or IV) is chosen for economic and other reasons. For these designs some of the interactions are aliased with the main effects or other interactions. To find a design to facilitate the estimation of the specified interactions, a traditional approach is to write down the strings of aliases (or the interaction table) and use trial and error to find a solution in which no two aliased effects are assigned to the specified interactions. Except for the well-trained, this process can be quite cumbersome. Taguchi (1959, 1987) proposed a method called linear graphs to solve this problem. (Incidentally the term "linear" is a misnomer. A literal translation of the original Japanese name should be dot-line graphs.)

Let me use the simplest example to explain his idea. For the 2^{4-1} design with $I = 1234$ as the defining relation, the six two-factor interactions are aliased in three pairs: $12 = 34$, $13 = 24$, $14 = 23$. If one interaction in each pair is insignificant, the other can be estimated. We can therefore estimate $\{12, 13, 14\}$ or $\{12, 13, 23\}$, which can be represented by the graphs in Figure 1. It is easy to show that these two graphs capture all of the solutions to the problem. The job of the experiment planner is simple. Draw a graph to depict the specified interactions, and compare it with a provided list of graphs to see if a matched graph can be found. In this case, it is easier for nonstatisticians to use the graph-aided method to solve the problem than to go through the algebra of aliasing relations.

As shown by Wu and Chen (1992), for larger problems Taguchi's method is deficient. For the 16-run 2^{n-p} designs, one of his graphs corresponds to a resolution V design for which it is not necessary to use graphs because all two-factor interactions are estimable. (It puzzles me that he was not aware of the

Figure 1. Linear Graphs for a Simple Example.

notion of resolution or its implication to the estimability of interactions.) The more serious problem is the use of resolution III designs. In general, there is no guarantee that the design represented by the graph has any good overall properties such as maximum resolution or minimum aberration. These problems are resolved in an improved version proposed by Wu and Chen (1992). For larger problems, however, any graph-aided method including the one proposed by Kacker and Tsui (1990) will become unwieldy. An alternative would be to use algorithms such as those of Franklin (1985).

Taguchi's main contribution is in the innovative use of graphs to capture solutions obtainable from the aliasing relations. For small to medium problems, the method of linear graphs and the modification due to Wu and Chen (1992) save experimenters from doing the tedious work of finding a feasible solution. Nonstatisticians are more willing to adopt the tool because of its simplicity and graphical appeal. Past experience has proved that user-friendly tools are more easily acceptable to the majority of our customers.

Anne Shoemaker and Kwok Tsui

Fractional factorial plans are commonly used in robust-design experiments, either as control or noise arrays or as "combined arrays" under the response-model approach to robust design (see Sec. 6). Maximum resolution (Box and Hunter 1961) and minimum aberration (Fries and Hunter 1980) are often used as criteria for planning fractional factorial experiments. These criteria basically assume that interactions of the same order are equally important and lower order interactions are more important than higher order interactions.

If physical knowledge suggests that certain interactions are likely to be important, however, we want a design that does not confound these interactions with each other. Maximum resolution and minimum aberration are not sufficient criteria to ensure this property in a design. Greenfield (1976) and Franklin and Bailey (1977) proposed a different criterion, which seeks a plan that allows the main effects and a specified set of interactions, called a "requirement set," to be estimable without being confounded with each other. There are situations in which prior knowledge about interactions is available and the requirement-set criterion conflicts with maximum resolution/ minimum aberration criteria. The experimenter then has to make a trade-off between these two criteria in planning the experiment.

Since these important design criteria generally may conflict with each other, new design optimization strategies are needed for planning industrial exper-
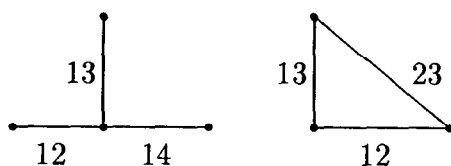
iments. One possible strategy, adopted by Wu and Chen (1992), is to search for designs that optimize resolution and aberration subject to the constraint that they must satisfy the requirement set. Alternatively, one could prioritize the importance of interactions in the requirement set, then drop the less important interactions to attain maximum resolution and minimum aberration.

George Box

Because engineers have traditionally relied on one-factor-at-a-time experimentation, main effects will often have already been put to use, and it will be the unexpected interaction that is waiting to be discovered and sometimes to be exploited with dramatic results (for example, see Box 1990; Hellstrand 1989).

Although occasionally it is possible to predict that certain factors are more likely than others to interact, predictions of this sort must be viewed with some skepticism. For example, it may be argued that factors occurring at different stages of a process will not interact. This is not always the case, however. For example, the best conditions for purifying a chemical may depend very much on the conditions used for its manufacture. A rather reckless extension of this idea is to say that a few expected interactions can be picked out from a much larger number of possible interactions and the remainder treated as inactive. It seems to be logically indefensible to say that we need an experiment to find out which factors have main effect (first-order effects) and at the same time claim that we know which factors have interactions (second-order effects). Whenever I work on planning an experiment and I draw diagrams to illustrate each possible two-factor interaction, I say to the experimenter, "Could something like this happen?" I almost always get the answer, "Yes, I can see how it could."

For me therefore, the most important rationale for the use of fractional designs is the separation of the "vital few (factors) from the trivial many" using the concept of design resolution. In addition, if we do want to isolate certain specific interactions, I fail to understand the supposed advantages and alleged simplicity claimed for Taguchi's linear graphs. The case of 8-run two-level designs is trivial and no help is needed. For 16-run designs, the graphs are complicated and even in their author's hands can produce designs that are demonstrably inferior (for example, see Box, in press) to those obtained by dropping or adding factors from designs of highest resolutions (resolution V with 5 factors, resolution IV with 8 factors, and resolution III with 15 factors).

6. Use of Combined Arrays and Direct Modeling of Response

There have been efforts at integrating Taguchi's parameter-design principles with well-established statistical techniques. Several authors have advocated treating noise factors (which are fixed during parameter-design experiments) as design factors, using a single design matrix, and modeling the response directly as a function of the control and noise factors. This section provides an overview of some of this work. See Easterling (1985) for an early reference to this approach.

Jerome Sacks and William Welch

The design and analysis strategy introduced by Welch, Yu, Kang, and Sacks (1990) and Yu, Kang, Sacks, and Welch (in press) implemented Taguchi's parameter-design objectives using more efficient, domestically developed techniques. We used response-surface methodology (RSM) to directly model the response as a function of control and noise factors. Our motivation was experiments conducted via computer simulation (about which we have more to say

in Sec. 7.3), but many of these ideas carry over to physical experiments.

As we see it, Taguchi's objectives can be simply formulated as follows. Let x and z be the vectors of control and noise factors, respectively, and let $y_i(x, z)$ denote the ith quality characteristic of interest ($i = 1, \ldots, q$). Given a loss function, $l[y_1(x, z), \ldots, y_q(x, z)]$, and, say, expected loss as a criterion, the objective is to find the value of x that minimizes

$$L(x) = \int l[y_1(x, z), \ldots, y_q(x, z)]f(z)dz, \quad (6.1)$$

where $f(z)$ is the probability density of z.

A succinct description of our implementation of this objective is as follows:

1.   Build a model for each $y_i(x, z)$ as a function of *all* factors, control, and noise.

2.   Replace $y_i(x, z)$ in the expected loss (6.1) by the fitted model $\hat{y}_i(x, z)$, and carry out the optimization via this cheap-to-compute surrogate.

Modeling the response, as opposed to an SN ratio, has several advantages. First, the single experimental array for both control and noise factors will usually require far fewer observations than Taguchi's crossed arrays, even when interactions between the control factors are included. Second, the engineer is more likely to have background knowledge when modeling the quality characteristic of interest than when modeling an SN ratio. Related to this, the model provides insight into how the factors affect the quality characteristic, a quantity of engineering relevance. Third, we have found that the quality characteristic is often easier to model than an SN ratio. A more accurate model leads to a more reliable optimization and ultimately a better engineering design; see the example of Welch et al. (1990).

Anne Shoemaker and Kwok Tsui

The response-model approach promises to be an effective framework for solving robust-design problems. The basic idea was first proposed by Welch et al. (1990). Related approaches were discussed by Shoemaker et al. (1991), Freeny and Nair (in press), and Montgomery (1991), in addition to the other discussants in this section.

The previous discussants have already mentioned some of the advantages of the response-model approach over Taguchi's approach. Other advantages were discussed by Shoemaker et al. (1991). As shown there, Taguchi's product-array format dictates estimation of all two-factor control-by-noise interactions, and often higher order "generalized" control-by-noise interactions as well. A combined array lets the experimenter choose the interactions to be estimated. This provides more flexibility so that the experimental budget can be used to fit models more refined than the main-effects-only models frequently used in the loss-model approach. Moreover, control-by-noise interactions provide special insights in the response-model approach because they are the effects that can be exploited to reduce response variability. In an IC-manufacturing example, Shoemaker et al. (1991) showed how examination of control-by-noise interaction plots reveals the mechanism by which two control factors dampen the effects of two noise factors. Finally, the wealth of techniques for empirical model building can be more easily applied to modeling the response than they can to the more specialized problem of modeling a variability measure.

Although the response-model approach is promising, the methodology for carrying it out is not yet mature. Since estimates of variance are based on the fitted response model, it is especially important that this model predict well. In addition, decisions on control-parameter settings can be very sensitive to how the response model is identified. Shoemaker et al. (1991) gave an example in which direct minimization of variance obtained from the fitted response model misses an important variability effect that was revealed by Taguchi-style analysis.

In this example, augmenting response-model analysis with examination of control-by-noise interaction plots revealed the missing variability effect. In general, however, analysis of control-by-noise plots may not lead to control-factor levels that minimize response variance. In a paper under preparation, we will show when these plots can fail and propose generalized control-by-noise plots and a data-analysis strategy that is more broadly useful.

The response-model approach requires a parsimonious model with good prediction capability. To attain this, it is important to use available physical knowledge. There are several ways that this might be done in the modeling process.

Choice of Response and Factors. As noted by some of the discussants in Section 3, physical knowledge should be used whenever possible to choose responses that are "fundamental." Failure to choose proper responses can induce nonlinearity, making it very difficult to find a well-fitting parsimonious model. Likewise, factors should be chosen in a way that simplifies their relationship with the response. The use of sliding factor levels has been noted by previous discussants. Phadke (1989, p. 145) gave an example of this for a photolithography process. The factors available to the engineer are aperture and exposure time, but the fundamental factors are depth of field and total energy. Phadke used sliding levels for exposure as a function of aperture to indirectly vary the fundamental factors and thus obtain a simple model. A further technique for simplifying models is transformation of response and factor variables.

Choice of Initial Model and Experiment Used to Estimate It. Sometimes enough physical knowledge may exist to suggest a specific response model, or a functional form for the response model. For example, in another IC-manufacturing application, Lin and Spanos (1990) had a theoretical model for polysilicon-deposition rate. To improve agreement between model-based predictions and empirical measurements, they used the functional form of the theoretical model but estimated the model's physical constants using a D-optimal experiment. This way, the theoretical model was "tuned" to a particular polysilicon-deposition machine and had very good prediction capability for that machine.

Identification of Fitted Model. Physical knowledge may resolve ambiguities induced by confounded effects and help identify a good model.

Raymond Myers and Geoffrey Vining

Several authors, including Vining and Myers (1990), have sought to combine Taguchi's parameter-design principles with conventional RMS. As the previous discussants have already noted, this approach incorporates the useful ideas in parameter design without suffering from the difficulties associated with Taguchi's methodology.

The approach postulates a single, formal model of the type

$$\hat{y} = f(\mathbf{x}, \mathbf{z}), \qquad (6.2)$$

where $\mathbf{x}$ and $\mathbf{z}$ represent the setting in the control and noise variables, respectively. It contains terms for both control and noise variables and all appropriate interactions. The noise variables are treated as fixed effects even though they are random effects in the process.

In addition to the advantages already pointed out by previous discussants, this approach allows one to provide separate estimates for the mean response and for variability rather than a single performance criterion; thus a variance response surface can be developed. Vining and Myers (1990) pointed out the natural link to the dual-response approach in RSM that many engineers find intuitively appealing. Compromise conditions between process mean and variability are easily visualized graphically. One can capture a sense of the process—that is, where in the space of the control factors the process is inconsistent and where the mean is desirable or unacceptable. Methodology developed by Myers and Carter (1973) can be used to generate graphics. In addition, existing software for doing graphical response-surface exploration with these two very natural responses can be useful.

We have received very favorable reaction from engineers and scientists to the notion of creating and exploring response-surface models for the response and variance. We will use an example to illustrate the ideas for variance modeling. Suppose we have control variables x_1 and x_2 and noise variables z_1 and z_2, with the specific fitted model in (6.2) given by

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2$$

$$+ b_{22} x_2^2 + c_1 z_1 + c_2 z_2 + \delta_{11} z_1 x_1 + \delta_{12} z_1 x_2. \qquad (6.3)$$

The significance of the two control-by-noise interactions carries important diagnostic information. Both control factors can be used to exert control on process variability produced by z_1. Note that z_2 does not interact with the control variables.

If the covariances among the noise variables are either known or well estimated, one can estimate the process variance by taking the appropriate variance operator in (6.3). At times, the noise variables may

be independent. If this is the case, the variance operator on Equation (6.3) gives (assuming z_1 and z_2 are scaled to $\pm\sigma$ and x_1 and x_2 to ± 1) $\hat{\sigma}_z^2(y) = c_2^2 + (c_1 + \delta_{11}x_1 + \delta_{22}x_2)^2 = c_2^2 + (\partial\hat{y}/\partial z_1)^2$. The choice of robust conditions implies choice of x_1 and x_2 that forces a "flat" $\partial\hat{y}/\partial z_1$. Suppose that $c_1 = \frac{1}{2}$, $\delta_{12} = \frac{1}{2}$, and $\delta_{12} = -\frac{1}{4}$. Movement away from the design center with x_1 in a negative direction and x_2 in a positive direction will result in a robust product. Consider Figure 2. We include the mean model obtained by an expectation operation on Equation (6.3) and the "line of minimum $\hat{\sigma}_z^2(y)$." For a target of $y = 50$, the optimum conditions are obvious. For the larger-the-better case, trade-offs and conditions for future experiments are evident.

In this illustration, we are, of course, assuming that the noise variables are independent in process conditions. When additional noise variables interact with control variables, the concept of a robust product resulting from 0 (or near 0) values of the slopes in the direction of $z_1, z_2, \ldots$, and so forth becomes apparent. In fact, the portion of the process variance that is influenced by the control variables becomes the squared length of the vector of these slopes. Similar interpretations surface when the z's are correlated.
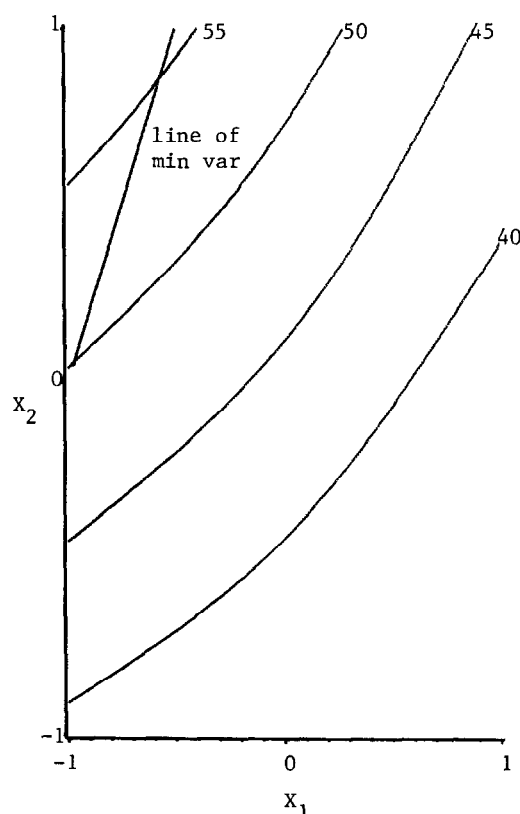

Figure 2. Dual Response-Surface Analysis.

George Box

The problem of robustness to environmental variables can be described as follows. We have a vector of design variables $\mathbf{x}$ that determines the design configuration for the product and a vector $\mathbf{z}$ that determines the environment conditions to which the product may be subjected within some practical region of variation R_z. Let y be some output characteristic such that

$$y = f(\mathbf{x}, \mathbf{z}) + e, \qquad (6.4)$$

where e is an experimental error and that ideally $E(y)$ should be equal to some value τ. Then, the environmental robustness problem is that of choosing a design configuration $\mathbf{x}_0$ so that in some sense $E(y)$ remains close to τ within the region R_z.

To solve such problems, Taguchi suggested the use of a cross-product experimental arrangement consisting of an "inner" or *design array* containing n design configurations and an "outer" or *environmental array* containing m environmental conditions. The environmental conditions could represent variation accidentally induced by the system, as in the well-known tile experiment of Taguchi and in the experiment described by Shoemaker et al. (1991). Alternatively, designated changes might be deliberately induced and suitably arranged in some specific experimental design.

Even when the design and environmental arrays are highly fractionated, the cross-product designs can result in rather large experimental runs so that the total amount of work required may be excessive. An important problem therefore is that of reducing the amount of effort needed in such a study. One approach, which was discussed in Section 2.2, is to select extreme conditions of the environmental noise variables (called compound variables in Sec. 2.2), and run only these. This method is always risky, however. When there are many environmental factors, guessing "extreme conditions" may be difficult or impossible. I will discuss another alternative of using split-plots in Section 7.1.

A third alternative, already noted by the previous discussants of this section, is to abandon the idea of the cross-product design and simply consider the design variables and the environmental variables $\mathbf{x}$ and $\mathbf{z}$ *together* as factors in a single design. Questions then arise as to what the structure of this experimental design should be. In particular, what "effects" of the design and environmental factors (linear, quadratic, interaction, etc.) it is important to estimate and how to choose experimental designs to achieve this.

One way to proceed (Box and Jones, in press) is (a) to represent $f(\mathbf{x},\mathbf{z})$, within specific ranges of the x's and the z's, by some suitable model (and in the

absence of more precise knowledge this might be a polynomial of degree d); (b) to formulate some measure of robustness such as $L(\mathbf{x}) = \int_{R_z} p(\mathbf{z})(y - \tau)^2 \, d\mathbf{z}$, the integrated square deviation from an ideal value τ, where in the absence of other knowledge $p(\mathbf{z})$ would be assumed to be uniform, and (c) to then consider what coefficients need to be estimated in $f(\mathbf{x}, \mathbf{z})$ to find the vector $\mathbf{x}_0$ that minimizes $L(\mathbf{x})$.

This approach leads to a number of interesting conclusions. For example, suppose that $f(\mathbf{x}, \mathbf{z})$ is represented by a second-degree polynomial, which, in an obvious matrix notation, may be written $f(\mathbf{x}, \mathbf{z}) = \beta_0 + \mathbf{x}'\beta + \mathbf{x}'\mathbf{Bx} + \mathbf{z}'\gamma + \mathbf{z}'\mathbf{Cz} + \mathbf{Z}'\mathbf{Dx}$. Then it can be shown under certain assumptions that to minimize $L(\mathbf{x})$ we do not need the individual elements of $\mathbf{C}$ but only tr $\mathbf{C}$. In other words, for the environmental factors we do not need to know their two factor interactions one with another; nor do we need to know the individual quadratic effects; but only their sum. Following this approach, the preceding authors obtained a number of designs for the quadratic model and for other models. These designs were considerably more economical than the cross-product designs.

James Lucas

For people who approach Taguchi with some background in RSM, the situation is especially simple. All of Taguchi's designs can be considered response-surface designs, and all response-surface tools are applicable. Taguchi's contributions, beyond the philosophical framework already mentioned, can be succinctly summarized as

1. Including environmental (noise variables) in your candidate list of design variables.
2. Using more screening designs. (For estimating first-order models, screening designs are two-level fractional factorials of resolution III.)

All the information about variability reduction and robustness can be extracted from the design using a response-surface analysis. Stationary points in the response surface, whether they be optima or saddle-points, will be conditions giving robustness. The propagation of error can be used to tie together the analysis proposed by Taguchi and traditional response-surface analysis (see Lucas 1991).

Thomas Lorenzen

In the following discussion, I will provide some comparisons of the different approaches to the design and analysis of robust-design experiments. This material is taken from a technical report by Lorenzen and Villalobos (1990) and represents joint work. A few new ideas are interspersed in the presentation.

As noted by the previous discussants, there are two basic approaches to designing a robust-design experiment, the product array and the combined array. The product array separately fractionates the control factors and the noise factors and forms the cross-product by running the noise array with every control factor combination in the control array. The combined array fractionates all of the factors.

General comparisons between the two designs are easy to make. The combined array usually has a better confounding pattern than the product array. On the other hand, more control-factor combinations are required in the combined-array design. If each control-factor combination requires building an expensive piece of hardware, the combined array can be more expensive. The analysis of the data for the product array is more intuitive because all noise combinations are comparable. The analysis for the combined array is not intuitive and requires estimating the missing noise combinations.

It is possible to compare the two approaches more formally in terms of detectability, the size of an effect that has, say, a 90% chance of significance. For the measured response, we want a design that has good detectability for control factors and "necessary" control-by-control interactions (for estimating the mean) and control-by-noise interactions (for estimating variability caused by the noise). Note that I said detectability, not just estimability, because the detectability of an eight-run experiment on six factors is over five σ, a worthless experiment because the effects have to be so large to be found.

Robustness measures such as the mean, variance, SN ratio, loss = squared deviation from target, loss after adjustment to target, and so on, however, are summaries *across* the noise combinations. The obvious and generally overlooked question is: How good is a design for robustness?

I will illustrate with a quick example, four control and two noise factors. The product array for estimating main effects only is a 2_{IV}^{4-1} design on the control factors crossed with a 2^2 design on the noise factors requiring 32 runs. The combined array for estimating all main effects and control-by-noise interactions is a 2_V^{6-1} design, also requiring 32 runs. Assuming all interactions except the eight control-by-noise interactions are negligible and can be pooled for error, both designs have *identical* detectability for main effects and interactions of the measured response variable.

For any robustness measure (summaries across noise combinations), the product array has eight observations (the eight control-factor combinations), but the combined array has 16 observations (all control-factor combinations are run). But, for the product array, four noise combinations are run with each

control combination, whereas only two, and not always the same two, noise combinations are run for each control combination in the combined array. In some sense, each robustness measure is only half as good in the combined array. How then does one compare the product and combined array for robustness?

Our solution is to formalize the "half as good" statement. For example, the number of data points used to estimate the mean of a main effect for a robust measure in the combined array is taken as four, not eight, since each observation is only half as good for the purpose of computing detectability.

Using this ad hoc procedure and still assuming that control-by-control interactions are negligible, the combined array has superior detectability for main effects—1.8 σ versus 2.5 σ. In addition, the assumptions can be checked in the combined array. Thus the combined array is superior for robustness.

We have software that computes detectability for the measured response and ad hoc detectability for robustness measures in a matter of seconds. This eliminates the need for general statements. But, for the dozen or so examples we have run, the combined array always had superior robustness properties.

There are three possible approaches for analyzing data from robust-design experiments. One is to compute and model directly the loss function or some other combined measure of the mean and standard deviation such as SN ratio or performance measure of independent adjustment (PerMIA) (see Leon et al. 1987). A second method is to separately model the mean and log-standard deviation and combine them to minimize loss. The third method is to model the raw data itself and predict all noise combinations for each control-factor combination. From this predicted data, the loss can be directly computed and the minimum selected.

Of the three methods, no guidelines currently exist for determining which procedure, if any, is preferred under general circumstances. Based on my limited experience, I would guess that modeling the raw data will turn out to be the best approach, followed by separately modeling the mean and log-standard deviation, and modeling the loss function directly will be the worst procedure, even though the loss function method is the easiest to teach and motivate.

The justification for guess 1 (modeling the raw data is superior to separately modeling the mean and log-standard deviation) is as follows. If the interaction between a control and noise variable is as in Figure 3a, then that control factor will not have an effect on either the mean or standard deviation calculated across the noise factor. The opportunity to improve the process by considering the midpoint of the two control levels is missed.
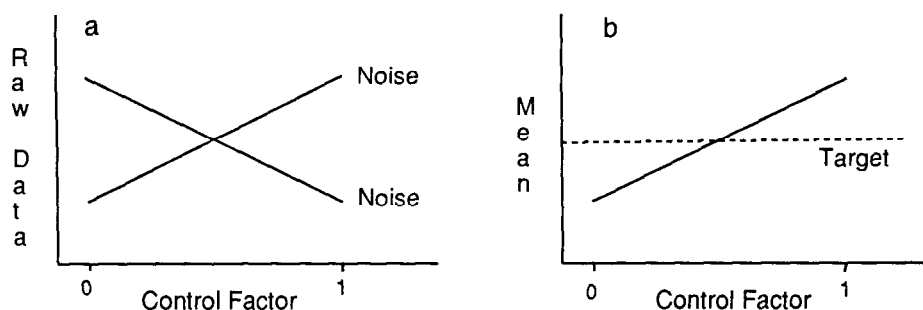
Figure 3. Simple Justification for Guesses 1 and 2.

The justification for guess 2 (modeling the mean and log-standard deviation is superior to modeling the loss function) is similar. If the effect of a control factor on the means is as in Figure 3b and that factor has no effect on the standard deviations, then that factor will have no apparent effect on the loss function. By modeling only the loss function, the opportunity to improve the process by considering the midpoint of the two levels is missed.

Within General Motors, engineers can design and analyze robust-design experiments through the use of an expert system called DEXPERT—Lorenzen and Truss (1990)—that essentially "black boxes" all difficult computations.

Raghu Kacker

The combined-array approach suggested by the previous discussants is just classical regression in which the explanatory variables are decomposed into control and noise factors. The constant-variance assumption that underlies this combined regression approach is unrealistic in many cases because not all sources of variation (noise factors) are likely to be included in the explanatory variables. The number

of significant sources of variation may be large and they may interact in complex ways to render the usual assumptions relating to the error distribution invalid. For a more general approach that allows for variance heterogeneity, see Freeny and Nair (in press).

Another assumption that underlies the combined regression approach is that the engineer can provide a prioritized list of the important control and noise-factor interactions. This is unlikely to be the case in most situations. In my collaborations with NIST scientists and engineers, I have not been able to get such a prioritized list in advance of the experiment. Even with experimental data in hand, engineers do not always agree on the priority list.

In experimentation, missing data and other disturbing outcomes often arise. The combined-regression approach is sensitive to missing data. Despite missing data, however, a product-array plan can usually provide information for further study. For example, two control-factor runs can be compared regardless of what befalls the measurements at the other runs. Similarly, control-factor effects can be determined with only one measurement for each run of the control array (see Liggett 1991).

7. Miscellaneous Topics

This section deals with a number of isolated topics— use of split-plot experiments, robustness to error transmission, computer experiments, use of GLM's for the joint modeling of mean and dispersion, Taguchi's and other techniques for analyzing nonstandard data, and dynamic parameter design problems.

7.1 THE USE OF SPLIT-PLOT EXPERIMENTS

George Box

When robustness experiments are carried out using cross-product designs, it is frequently most convenient to conduct them in a split-plot mode. In particular, examples described by Phadke et al. (1983),

Quinlan (1985), and Shoemaker et al. (1991) are clearly of this type. As is well known, misleading results may be produced by failing to take account of split-plot structure. For example, Quinlan (1985) used a saturated 16-run design containing 15 factors and tested four sample pieces of the cable from each run. In an analysis by Box (1988), the estimated variances for the between-run error were 13 times the size of the within-run error. Failure to properly account for these error sources may account for the plethora of significant effects found by Nelder and Lee (1991) in their reanalysis of the Quinlan data (see Sec. 7.4).

The concept of designing products that were robust

to environmental factors and the value of split-plot experiments in achieving this was well understood almost three decades ago by Michaels (1964). He described these ideas in the example from detergent testing at Proctor and Gamble Ltd. in the United Kingdom. In particular, he said:

Environmental factors, such as water hardness and washing techniques, are included in the experiment because we want to know if our products perform equally well vis-a-vis competition in all environments. In other words, we want to know if there are any Product × Environment interactions. Main effects of environmental factors, on the other hand, are not particularly important to us. These treatments are therefore applied to the Main Plots, and are hence not estimated as precisely as the Sub-plot treatments and their interactions. The test products are of course applied to the Sub-plots. (p. 223)

Conducting designs in split-plot mode does not, of course, change the number $m \times n$ of cells in the design, but it does change the structure of the error term e in Equation (6.4) and can greatly change the amount of experimental effort required. Box and Jones (1992) discussed a cake-mix example with a design (inner) array with $n = 9$ runs and an environmental (outer) array with $m = 5$ runs. Hence the cross-product design will have $m \times n = 45$ runs. As an illustration, Table 1 gives a list of alternative designs for this example. A fuller discussion of such arrangements and their usefulness and analysis in the context of robust design will be found in the work of Box and Jones (1992). In a valuable discussion of split-plot designs, Cox (1958) characterized whole-plot factors as "classification" factors. Although, as Michaels says, the preferred split-plot design would normally have the environmental variables as classification factors (design 3), sometimes experimental conditions and the relative cost of the various operations may point to a different arrangement.

In experimental solutions similar to the cake-mix example, a very attractive alternative, both to a full randomized design and to a split-plot design, is the *strip-block* experimental design (design 4). In the cake-mix example, a single replicate would merely

Table 1. Comparison of Designs

Type	Number of design configurations prepared	Number of environmental conditions	Number of experimental operations
1. Fully randomized design	$n \times m$ (45 mixes)	$n \times m$ (45 bakings)	$2nm$ (90)
2. Design variables = whole plots	n (9 mixes)	$n \times m$ (45 bakings)	$n(m + 1)$ (54)
3. Environmental variables = whole plots	nm (45 mixes)	m (5 bakings)	$m(n + 1)$ (50)
4. Strip block	n (9 mixes)	m (5 bakings)	$n + m$ (14)

involve the preparation of nine cake mixes, each of which could be divided into five parts and tested only in five bakings; thus a number of replicates of the design might be run with no more effort than would be required for a single replicate of an alternative design.

7.2 ROBUSTNESS TO ERROR TRANSMISSION

George Box

Taguchi discussed the problem of robustness to error transmission in which the exact mathematical relation $y = f(\mathbf{x})$ between the quality characteristic y and its components $\mathbf{x}$ is *known*. For example, in the design of an assembly, such as an electrical circuit, the relationship between the output voltage y of the circuit and the components (resistors, capacitors, etc.) may be known from physics. Variation in component characteristics around their nominal values is transmitted as variation in the response. There may be an infinite variety of configurations of $\mathbf{x}$ that can give a working assembly that can produce a desired mean value $E(y)$. An opportunity therefore exists to choose a configuration that is least affected by variation in the components.

Suppose the characteristics $\mathbf{x}$ of the components vary about "nominal values" ξ with *known* covariance matrix V. Thus, for example, a particular resistance x_i might vary about its nominal value ξ_i with known variance σ_i^2. (Moreover, variation in one component would usually be independent of that of another so that V would be diagonal.) Now variation in the input characteristics $\mathbf{x}$ will transmit variation to the quality characteristic y. Let us denote by $v(y)$ some measure of this transmitted variation. This could, for example, be the transmitted variance $\sigma^2(y)$ itself or some other measure such as $\log \sigma^2(\log y)$ or, almost equivalently, Taguchi's signal-to-noise ratio SN.

Using a Wheatstone-bridge circuit for illustration, Taguchi and Wu (1985) posed the problem of choosing ξ so that $v(y)$ is minimized. To solve it, they again employed an experimental design strategy using inner and outer arrays. Box and Fung (1986) pointed out, however, that since $f(\mathbf{x})$ is assumed known, $v(y)$ is a function of ξ that can be computed by well-known error transmission formulas and minimized using a standard optimization program or equivalently by response-surface methods. For an early example of using error-transmission formulas for the study of variability in engineering designs, see Morrison (1957), who gave a fully worked out example of this approach and who remarked:

Most engineering design is based to a large extent on relations between mean values or target values of the design parameters. For a statistical formulation, each equation of mean values should be supplemented by an equation in terms of the variance of the design parameters. This will require a knowledge of the component variances, which should be based on actual measurements rather on guesswork. (p. 133)

The warning in the last sentence is particularly apt because procedures for choosing a robust product can be very unrobust from a statistical point of view if, as is usually the case, a particular error structure must be assumed to apply over wide ranges of the ξ_i's. In the Wheatstone-bridge experiment, for example, Taguchi and Wu (1985) tacitly assumed that σ_i is proportional to ξ_i for each of the component characteristics and that this proportionality applies over a 25-to-1 range of variation of ξ_i. If we suppose more generally that $\sigma_i = g(\xi_i)$, the solution can be extremely sensitive to the choice of $g(\cdot)$.

As a simple illustration (for example, see Fung 1986), consider the problem of choosing a pendulum whose length x may be in error and that we wish to choose the minimal value of ξ of the length so that the *percentage error* in the period y of the pendulum is minimized. Suppose, as would be approximately true, that $y = cx^{1/2}$, where c is a constant; then it is easy to show that (a) if $\sigma(x)$ is independent of x, the *longest* pendulum possible should be used; (b) if $\sigma(x)$ is proportional to ξ, it *makes no difference* what the length of the pendulum is; and (c) if σ is proportional to ξ^α, with α greater than 1, then we should use the *shortest* possible pendulum. Furthermore, although for an electrical circuit it is reasonable to assume that the relation $y = f(x)$ is known, when, as is more usually the case, it must be estimated experimentally the problems are much more complicated and require further study.

7.3 COMPUTER EXPERIMENTS

Jerome Sacks and William Welch

Many parameter-design experiments are now run via computer models (CAD/CAM tools). This is particularly true in the design of integrated circuits, the area in which we have the most firsthand experience and in which we shall have most to say. Finite-element applications in mechanical engineering are also common, however. For instance, we were peripherally involved with a project to design a truck's brake caliper via computer simulation. There were about 40 parameters—dimensions, angles, and so forth— and several quality characteristics, the most important of which was the caliper's deflection.

In IC applications, the use of computer experiments to reduce the impact of noise variables has long been recognized. When Brayton, Hachtel, and Sangiovanni-Vincentelli (1981) surveyed circuit-

optimization techniques, Taguchi's parameter design was still largely unknown in North America. The techniques suggested for robust design are quite different from those of Taguchi; most notably, there is much emphasis in the electrical-engineering literature on the trade-off between multiple objectives and constraints.

The formulation we described in Section 6 leading to the minimization of the expected loss in (6.1) or some similar criterion is particularly suited to computer experiments. All noise factors are identified and easily manipulated. The computer codes are typically deterministic, in which case there is no random-error term representing unmodeled factors. In a physical experiment, if random error of important size is present, then any functional dependence of the error variance on the control factors would have to be modeled in our formulation. There are other distinctions between data from deterministic computer models and those from physical experiments. For example, complex relationships can be uncovered with far fewer observations when random error is absent.

Taguchi used a deterministic mathematical model to generate data in his Wheatstone-bridge example (Taguchi 1986, chap. 6). There, the mathematical equation is trivial, so it is not clear why one would not simply plug the objective function into a numerical optimizer. Admittedly, this might produce a local optimum, but Taguchi's solution is also suboptimal (Box and Fung 1986). Real computer models can be computationally very expensive, and direct numerical optimization [or the variants surveyed by Brayton et al. (1981)] can require too many function evaluations. Similarly, Taguchi's experimental plan with 1,296 observations is too expensive for real applications. Thus Taguchi's approach appears (to us anyway) to be overcomplicated for simple deterministic models and too expensive for realistic problems. The strategy we outline has successfully tackled problems much more complex than the Wheatstone-bridge example with far fewer observations.

To illustrate the magnitude and complexity of the real problems that engineers are trying to tackle via computer experiments, we will give a brief overview of an ongoing project to design an IC. There are more than 10 quality responses of interest, including two primary and two secondary time delays, four power supply "peaks," two output impedances, and two output currents. These responses depend on 20 controllable device (transistor) sizes and 16 noise factors representing variations in the manufacturing-process conditions. The noise factors have known normal distributions, based on empirical measurements of the manufacturing process. Eight of them are correlated. This computer model is moderately

expensive to compute—100 runs of the model take approximately 24 hours on a workstation.

Loosely stated, the engineering objectives were given in terms of designing for the worst case: Find the device sizes that minimize the maximum of the four power supply peaks, subject to constraints on the remaining responses (e.g., an upper bound on each primary delay). In the presence of the noise variations, an upper bound on, say, a delay time translates into an upper bound on the mean plus three standard deviations. Like the smaller-the-better SN ratio, this penalizes both the mean and the standard deviation, but it relates directly to the engineering criteria. Similarly, a power-supply peak is the mean plus three standard deviations. These worst-case criteria cannot be written down as an expected loss (6.1), but this demonstrates the flexibility of our implementation. We always model and predict the basic quality characteristics. During optimization via these predictors, the engineer can specify an objective function for the particular problem. In our experience, multiple conflicting quality characteristics necessitate some customization of the objective. Portmanteau criteria are too inflexible.

A major complication in this problem is that the various end uses for the circuit call for different values of the time delay upper bound. By building cheap-to-compute approximations for all response functions, an engineer can substitute the particular delay time in the objective function. For reliable optimization, high accuracy of prediction is necessary, and specifications for some responses were provided.

This problem is clearly of considerable magnitude and complexity. There are toy problems and real problems: We find the latter more helpful for directing research. Because physical experiments of this size are infeasible and computer experiments are relatively new (to the statistical community anyway), little attention has been paid to complex problems in the literature. The details of our analysis of this example will appear elsewhere as a case study.

As described in Section 6, our approach builds approximating functions relating each response to all input parameters—control and noise—and then optimizes via these approximations rather than directly through the computationally expensive computer model. The reliability of the optimization clearly depends on the accuracies of the approximating models. Sometimes, when factor ranges are sufficiently narrow, we have found second-order polynomial models to give enough accuracy. When the factors have wide ranges, however, leading to complex input-output relationships, or when data are scarce, the interpolators described by Currin, Mitchell, Morris, and Ylvisaker (1991), Sacks, Schiller, and Welch (1989), Sacks, Welch, Mitchell, and Wynn (1989), and Welch

et al. (1992) are more flexible and data-adaptive and tend to be more accurate and successful for prediction and optimization.

Computational cost in running the simulator often dictates comparatively small experimental designs to fit the approximating functions. We have found Latin hypercube designs (McKay, Conover, and Beckman 1979) useful and easy to construct. Moreover, they can incorporate correlations between the noise factors when they exist (Iman and Conover 1982). With many factors over wide ranges and optimization as an objective, it is too much to expect that a single-stage experiment will be economic or effective. The sequential strategy of Bernardo et al. (in press) took two stages in one example with 14 factors and four quality characteristics. The information from the first stage was used to zero in on a subregion for which accurate prediction and reliable optimization was possible. The two Latin hypercubes required a total of just 125 observations. We do not see how a Taguchi-style experimental plan crossing two orthogonal arrays could achieve the same goals without far more experimental effort.

A fuller summary of this work on quality improvement via computer experiments can be found in the work of Welch and Sacks (1991).

For the strategy we outlined for parameter design via computer experiments, sophisticated software tools are essential. Computer-intensive methods like the stochastic-process interpolators of Sacks, Welch, Mitchell, and Wynn (1989) extract maximum accuracy of prediction from costly data. Other computer-intensive function-fitting algorithms such as multivariate adaptive regressive splines (Friedman 1991) also may have utility. We are currently developing the software we have written for our research purposes into a system suitable for wider dissemination. It will enable engineers to initially identify the important factors, build approximating models, visualize the input-output relationships, and proceed sequentially to a good engineering design.

7.4 GENERALIZED LINEAR MODELS FOR THE JOINT MODELING OF MEAN AND DISPERSION

J. A. Nelder

I will begin with a brief introduction to GLM's.

An Outline of GLM's. GLM's extend the class of classical linear models in two ways. First, they allow the errors to come from a class of distributions instead of just the normal distribution. This class (known to statisticians as one-parameter exponential) includes, as well as the normal, the Poisson, binomial, multinomial, gamma, negative-binomial, and inverse Gaussian distributions.

Second, a GLM splits the finding of an additive scale for the effects of the explanatory variables from the specification of the error structure. The scale on which the effects are assumed additive is related to the mean of the error distribution by the link function. Thus we write $\eta = \Sigma \beta_j x_j$ for the linear part of the model, where η is called the linear predictor, and connect this with the mean μ by the link function $\eta = g(\mu)$. We do not transform the data to produce additivity; rather we transform the hypothetical mean values. For example, the log-linear model, which is used for the analysis of counts, is a GLM in which the error distribution is Poisson and the link function is the log.

Many characteristics of classical linear models generalize immediately to GLM's. These include the structure of the linear predictor, the ANOVA table of a nested set of models, and model-checking ideas like residuals, leverage, influence, and so forth. Furthermore, a single algorithm, a version of iterative weighted least squares, fits all GLM's. See McCullagh and Nelder (1989) for a full treatment of these models.

A very important property of all GLM's is the form of the variance var$(y) = \phi V(\mu)$. This shows that the variance splits into two parts—ϕ, called the dispersion parameters, which is independent of the mean, and $V(\mu)$, called the variance function, which describes how the variance changes with the mean. In the terminology of Leon et al. (1987), ϕ is a PerMIA. Box's (1988) criterion of separation can be stated for GLM's as finding the appropriate variance function for the data. Similarly, modeling the variance is generalized to modeling the dispersion; the two are the same only for normal errors. Box's (1988) second criterion, parsimony, is interpreted as finding an appropriate link function to produce additivity of effects, together with a parsimonious set of explanatory variables that accounts well for the variability in the response.

GLM's for Parameter-Design Experiments. The aim in most parameter-design experiments is to reduce the variation in the products/processes while holding the mean at the target value. Thus we require designs supporting the joint modeling of both mean and dispersion. To do this, we use a pair of interlinked GLM's, one for the mean and the other for the dispersion. Each has a response variable, a variance function describing how the variance depends on the mean, a link function defining a scale on which the effects of the explanatory variables are assumed additive, and a set of explanatory variables contributing to the linear predictor.

In Table 2, the response variable for dispersion, d, is the deviance component [a generalization of the

Table 2. GLM's for Joint Modeling of Mean and Dispersion

Components	Mean	Dispersion
Response variable	y	d
Mean	μ	ϕ
Variance function	$V(\mu)$	gamma
Link function	$\eta = g(\mu)$	$\zeta = h(\phi)$
Linear predictor	$\eta = \Sigma \beta_j x_j$	$\zeta = \Sigma \gamma_k u_k$

squared residual $(y - \hat{\mu})^2$ for normal errors] associated with the GLM for the mean. The justification for fixing the variance function for the dispersion as gamma—that is, ϕ^2—is that the deviance has a distribution close to the gamma even when the error of y is not normal. For parameter-design experiments, the explanatory variables used in the linear predictor for the mean and dispersion are both given by the configurations of the control factors in the design matrix. We use separate notations (x_j and u_k) to emphasize that the factors that are important for dispersion may or may not occur in the model for the mean. A term occurring in the mean linear predictor only can thus be used to get the mean close to target, while a term in the dispersion linear predictor, whether or not it occurs also in the mean, can be used to reduce dispersion. It is common to find that the link function $h(\cdot)$ for the dispersion can be taken as the log.

The fitting of this model uses as an optimizing criterion the idea of extended quasi-likelihood first defined by Nelder and Pregibon (1987), further developed by McCullagh and Nelder (1989), and exemplified by Nelder and Lee (1991). The algorithm is an extension of the standard GLM algorithm, in which the GLM for the mean is fitted, assuming that the fitted values for the dispersion are known, and that for the dispersion is fitted using the fitted values for the mean to form the response variable d. The fitting alternates between the mean and dispersion models until convergence is achieved. GLIM macros for fitting these models are available from the author.

A Strategy for Fitting. We first seek separation, which is here interpreted as finding a suitable variance function for the mean. We fit saturated models for both the mean and dispersion, using variance functions from a family (say, the Box–Cox power family), and search for a minimum of the extended quasi-likelihood. We now seek parsimony, looking for link functions and a parsimonious set of terms in the explanatory variables for both mean and dispersion. We begin with a saturated model for the mean and analyze the dispersion. Then, using the weights derived from the reciprocals of the fitted dispersions, we model the mean.

The next step is to check the two models for in-

ternal consistency (see McCullagh and Nelder 1989, chap. 12), going back over the previous steps if necessary. When the checks are satisfactory, we can proceed to prediction from the models by finding optimum settings of the explanatory variables for the purpose in hand.

An Example. Nelder and Lee (1991) applied this technique to the data set on the shrinkage of speedometer cables (see Box 1988). The design was a saturated fractional factorial with 15 factors labeled $A-O$, each at two levels, with four samples per run. The response variable is a ratio of continuous variables so that the data require a variance function that tends to 0 with the mean. We used the family $\mu^\theta(1 - \mu)^\theta$ and found $\theta = 1$ to be satisfactory. The effects of the two most important factors, E and G, on the mean were multiplicative, showing that a log link function was needed for the mean. The analysis of the dispersion, using a log link and gamma errors, gave a model with five factors, D F G H N. This predicted well two within-run extreme variances, one high and one low.

This data set was also analyzed by Box (1988) using data transformations. For the mean, Box's analysis identified only factors E and G, where Quinlan's original analysis used eight factors, A C D E F G H K. Our analysis indicated a case for two additional factors L and N, giving 10 in all. For this experiment, there were two independent confirmatory runs (called Before and After) against which model predictions can be checked. For the run Before, the mean model with 10 factors was particularly successful in predicting the mean, though less so for the run After, although here the predictions of the models with 8 or 10 factors were considerably better than the model with only two.

Conclusion. The model class described previously is general enough to cover the analysis of parameter-design experiments in which the response is continuous, or is a count of proportions, and to do so in a unified way; it allows the description of separation and parsimony quite independently and can be fitted by a small extension of a standard algorithm.

7.5 ANALYSIS OF NONSTANDARD DATA

Jeff Wu

Nonstandard responses such as binary (good or defective), ordered categorical (window not open, small, medium, large), Poisson (number of defective chips on a wafer), or censored (typical in life testing) are quite common in experimental situations. Standard textbooks on experimental design do not give special attention to these problems. A common approach is to transform the data to near normality and

then use the wealth of tools for analyzing normal data. Near normality cannot be achieved, for example, when the data are sparse and the response is binomial or censored. A direct approach would be to model the response by an appropriate likelihood and use standard methods for estimation. It has at least two problems. First, the maximum likelihood estimates (or estimates based on other likelihood related methods) do not often exist when there are strong factorial effects. See Hamada and Tse (in press) and references therein for precise conditions. Second, there are too many models to be entertained because of effects aliasing. Both points were discussed by Hamada and Wu (1991) in the context of censored data.

Perhaps recognizing the limitations of the methods available to him, Taguchi proposed the *accumulation analysis* for analyzing ordered categorical data and the *minute acumulating analysis* for analyzing censored data. A general conclusion based on the work of Nair (1986), Box and Jones (1986), and Hamada and Wu (1990) on the former and the work of Hamada (in press) on the latter is that they are unnecessarily complicated, inefficient, or even invalid. Although the method of scorings is simple and can be effective for a certain type of ordered categorical data, it is still a challenging problem to find a sound method of analysis when the replicates are few per run, the design matrix is sparse, and the number of categories is only two or three. For censored data, more sound methods have been proposed, but they still experience some problems when the maximum likelihood estimator does not exist (see Hamada and Wu [1991] and references therein). Hamada and I are working on a Bayesian modification of our method.

Although Taguchi's analysis methods are faulty, he deserves credit for bringing to our attention this class of problems and for encouraging the collection and analysis of such data in industry as evidenced by the many case studies he and his colleagues have presented. In particular, his emphasis on using highly fractionated experiments to increase lifetime or improve reliability is a valuable addition to the reliability literature, which tends to be more interested in estimating than improving reliability. Because of the technical difficulties associated with fractionated experiments, there is a great opportunity for research in this area. One possibility is to modify GLM for these and other types of nonstandard data.

7.6 DYNAMIC PARAMETER-DESIGN PROBLEMS

Shin Taguchi

As I stated in Section 3, whenever possible dynamic characteristics should be used in parameter-

design applications. This is a much more powerful approach than treating the problems as static problems.

In the coating-process application I discussed in Section 3, we can treat the response (thickness) as a static problem and use the corresponding SN ratio to achieve robust performance. It is more powerful, however, to treat this as a dynamic problem. The response y = thickness is related to M = spray time; ideally, y should be proportional to M. Variability around this proportional relation due to noise factors creates problems such as voids, orange peel, poor appearance, and low yield. Therefore, we want to minimize this variability.

Here y is the output response, and M is called a signal factor. A signal factor is an input to an engineering system. This is an example of a dynamic problem in which the ideal response should track the signal. The SN ratio for this problem is

$$SN = 10 \log_{10}\beta^2/\sigma^2, \qquad (7.1)$$

where β measures the slope of the linear relationship forced to go through 0—that is, $y = \beta M$—and σ^2 is the mean squared deviation due to all other sources such as noise effects and nonlinearity. The optimization problem follows a two-step process just as in a nominal-the-best case:

1. Find control factors to maximize SN.
2. Adjust β to the desired sensitivity level.

The second step is called a leveling or sensitivity adjustment, which is essentially the same as a tuning activity.

Dynamic characteristics are being used increasingly in parameter-design applications. In fact, more than half of the case studies presented in the latest Taguchi symposia in the United States and Japan involved dynamic characteristics. The training by the American Supplier Institute has been changed greatly to reflect this emphasis.

Raghu Kacker

Dynamic systems are characterized by the presence of a signal factor. Taguchi's most well-known SN ratio for dynamic problems, given by (7.1), arises naturally in metrology, where the property of interest is the signal factor and the property actually measured is the response. Indeed, except for the log transformation, this criterion was introduced under the name "sensitivity" by Mandel and Stiehler (1954). In this, the criterion is not new.

Taguchi estimates (7.1) from the mean squares of the ANOVA table for a simple linear regression. This is a monotonic function of both the F ratio and the coefficient of determination R^2. Thus these are

all equivalent performance statistics. Of course, from a data-analytic point of view, one of them (or its transformed version) may be preferable.

When many seemingly equivalent performance statistics are available, a choice must be made. The distributional properties of the chosen statistic is an important consideration for the following reason. When a performance statistic is to be used to compare two or more systems, two questions naturally arise: Is the difference between the systems significant and is the performance statistic sufficiently large? These questions can be addressed when the distribution of the performance statistic is known under various hypotheses. I am addressing these issues in a forthcoming paper.

Jeff Wu

It is only recently that Taguchi and his colleagues at the American Supplier Institute have started promoting the importance of problems related to "dynamic characteristics." As far as I know, little work has appeared in the statistical literature on the kind of problems that Taguchi and his colleagues at the National Research Laboratory of Metrology (Tsukuba, Japan) have worked on. Since no clear definition has been given by Taguchi, I will give our (Miller and Wu 1991) definition as follows. A dynamic system can be described by the schematic diagram in Figure 4.

If the input signal has only one level, it is called a static system. When the input signal is used to control the response signal, it is called a dynamic system. Since several levels of the input signal are to be entertained, one should make the system efficient over a range of the input signal. Examples include calibration of a measurement system, injection molding process, and steering mechanism of a car. The term "dynamic" may be misleading because it does not properly describe the measurement-system problem. A better description is through a common feature of the three problems; that is, the response is a functional relation between two quantities, and statistical modeling is to be done on this relationship. In the case of the steering mechanism, feedback control is not incorporated in either Taguchi's or our formulation.

Let me now turn my attention to the problem of improving a calibration system through a parameter-
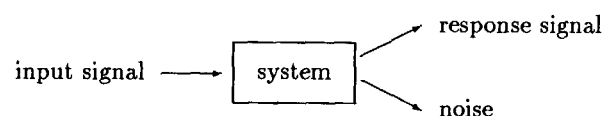

Figure 4. Schematic Diagram for a Dynamic System.

design experiment. In addition to its practical importance, it is interesting to note that in this case, Taguchi's SN ratio (7.1) is a sound choice. Formally, the system can be described by $y = \alpha + \beta U + \varepsilon$, $\text{var}(\varepsilon) = \sigma^2$, where y is a measurement of W and W is related to U, the unknown quantity of interest, through $W = \alpha + \beta U$. The slope β and the error variance σ^2 depend on some control factors. The purpose of parameter design is to choose the control-factor settings so that the quantity of interest $u_0 = (y_0 - \alpha)/\beta$, where y_0 is the measured value of W, can be estimated accurately. Taguchi (1987, chap. 22) showed that, when α and β are assumed known, the mean squared error of u_0 is minimized by maximizing the SN ratio β^2/σ^2 or equivalently $\log(\beta^2/\sigma^2)$ [see (7.1)]. A more sound and rigorous approach is to express the purpose as minimizing the length of the Fieller (1954) interval, which is known to be exact for estimating u_0 in inverse regression. Miller and Wu (1991) showed that the length of the Fieller interval is a decreasing function in the SN ratio β^2/σ^2, thereby justifying the choice. Taguchi's analysis strategy is to model $\log(\hat{\beta}^2/s^2)$ as a function of the control factors, where $\hat{\beta}$ and s^2 are, respectively, the

least squares estimate of the slope and the variance estimate of σ^2 for each control run. This modeling technique shares the same problem as the SN ratios for static problems. Although the SN ratio is a performance measure to be maximized in this case, it is not always easy to model it directly in terms of the control factors. It is better to separate performance measure maximization from statistical modeling. Miller and Wu (1991) proposed a response-function model consisting of $y_{ij} = \alpha_i + \beta_i u_j + \sigma_i \varepsilon_{ij}$, $\beta_i = X_i \theta_\beta + \sigma_\beta \tau_i$, and $\log \sigma_i^2 = X_i \theta_\sigma + \sigma_\sigma \xi_i$, where i denotes the ith control run, X_i the ith control-factor setting, $\{u_j\}$ is a collection of known quantities of U for the purpose of calibration, and $\text{var}(\varepsilon_{ij}) = \text{var}(\tau_i) = \text{var}(\xi_i) = 1$. It enables the investigator to study which factors affect β and which factors affect σ^2. By combining the three equations, it allows direct modeling of the response y_{ij} as a function of X_i. The functional relationship between y and u can also be studied. In contrast, Taguchi's performance-measure modeling compresses the data y_{ij} into a single measure and may result in loss of information. Details including reanalysis of Taguchi's drive-shaft data can be found in the work of Miller and Wu (1991).

8. Some Concluding Remarks

Bovas Abraham and Jock MacKay

It is our experience that the key to success in using variation-reduction experiments is the use of a systematic approach. This means that great care must be given to defining the problem, assuring good measurement systems, identifying the noise factors causing the problem, learning the behavior of the noise factors, choosing the control factors and their levels, selecting the design given the production constraints, analyzing and presenting the results, drawing and confirming the conclusions, and standardizing the recommendations. There are many opportunities for the statistician to exhibit technical skills, but a much more important role is to ensure that the systematic approach is followed. This point should be remembered both when teaching and consulting.

In many instances, the design and ensuing analysis are very simple due to understanding of the problem and production constraints. The availability of elaborate software is not often necessary. Success depends instead on rigorously following a systematic approach with a team of highly knowledgeable production people. The goal of the statistician should be to bring together the process knowledge, the disciplined approach, and the appropriate statistical tools.

Anne Shoemaker and Kwok Tsui

Although robust-design experimentation methods are now covered in some university quality-control and design-of-experiments courses, the primary vehicle for teaching robust-design methods to engineers is industrial short courses. Since different solution methods are appropriate for different application areas, it is crucial that these short courses focus on one homogeneous audience and present methods that are appropriate to their application and are easily integrated with their work processes.

The training should start with the problem of robust design and then present a step-by-step solution procedure, illustrated with examples from the student's work area. Methods should be taught only at the points in the procedure where they are used. Intuitive justification of methods is preferable to theory.

Software can increase the effectiveness of training by allowing students in-class hands-on experience and giving them something to take away that mirrors the step-by-step solution they learned in class. At AT&T, we have used Robust Design Experimenter, a personal computer software system that has an interface so simple it has virtually no learning curve. This sys-

tem also removes major bottlenecks to technology transfer by providing an "automatic experiment planner" (Tsui 1989) that can construct mixed-level fractional factorial experiments from a list of required main effects and interactions, using very simple analysis methods such as main-effect and interaction plots and a "trade-off table" for evaluating several responses simultaneously. Additional graphical tools such as half-normal probability plots (Daniel 1976) are included in other software systems.

When software is not available, simple graphical and tabular tools should be taught to help engineers plan robust-design experiments by themselves. Interaction graphs (Kacker and Tsui 1990), improved linear graphs (Wu and Chen 1992), and confounding tables (Tsui 1988) are effective for this purpose.

Raymond Myers and Geoffrey Vining

Taguchi has helped draw considerable attention to benefits of statistical methods in industry. Many of his ideas in parameter design will continue to motivate activity by the user and statistical researcher. The influence has been more widespread than many think. A nice foundation is in place for mean and variance modeling. It is not our intention to imply that this is the only form of analysis of the data. Indeed, much work lies ahead in the development of analytical techniques and experimental designs. We must remember, however, that it will take time before parameter design is adopted at the level that professional statisticians would like. Many potential practitioners have not begun.

We tend to generalize about parameter-design usage because of information that reflects our own experience. We tend to forget that *only a small percent of American companies use statistical methods at all.* One thing is certain—we hope that a survey 10 years hence will reveal a profound increase in usage, with the usage involving efficient methodology. There is much more communication to be done at a lower level. In a recent quality symposium, George Box indicated that he would be happy if all engineers would merely design a 2^3 factorial experiment. Sadly, only a small portion have.

ACKNOWLEDGMENTS

REFERENCES

AT&T (1956), *Statistical Quality Control Handbook*, Indianapolis: Author.

Bartlett, M. S., and Kendall, D. G. (1946), "The Statistical Analysis of Variance Heterogeneity and the Logarithmic Transformation," *Journal of the Royal Statistical Society*, Ser. B, 8, 128–150.

Beckhofer, R. E. (1960), "A Multiplicative Model for Analyzing Variances Which Are Affected by Several Factors," *Journal of the American Statistical Association*, 55, 245–264.

Bernardo, M. C., Buck, R., Liu, L., Nazaret, W. A., Sacks, J., and Welch, W. J. (in press), "Integrated Circuit Design Optimization Using a Sequential Strategy," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 11.

Bisgaard, S., and Diamond, N. (1991), "An Analysis of Taguchi's Method of Confirmatory Trials," Report 60, University of Wisconsin-Madison, Center for Quality and Productivity Improvement.

Bowman, K. O., Hopp, T. H., Kacker, R. N., and Lundegard, R. J. (1991), "Statistical Quality Control Techniques in Japan," *Chance*, 4, 15–21.

Box, G. E. P. (1952), "Multi-Factor Designs of First Order," *Biometrika*, 39, 49–67.

—— (1988), "Signal-to-Noise Ratios, Performance Criteria, and Transformations" (with discussion), *Technometrics*, 30, 1–40.

—— (1990), "Do Interactions Matter?" *Quality Engineering*, 2, 365–369.

—— (in press), "What Can You Find Out With Sixteen Experimental Runs?" *Quality Engineering*, 5.

Box, G. E. P., and Fung, C. A. (1986), "Studies in Quality Improvement: Minimizing Transmitted Variation by Parameter Design," Report 8, University of Wisconsin-Madison, Center for Quality and Productivity Improvement.

Box, G. E. P., and Hunter, J. S. (1961), "The $2^{(k-p)}$ Fractional Factorial Designs," *Technometrics*, 3, 311.

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

Box, G. E. P., and Jones, S. (1986), Discussion of "Testing in Industrial Experiments With Ordered Categorical Data," by Vijayan N. Nair, *Technometrics*, 28, 295–301.

—— (1990), "An Investigation of the Method of Accumulation Analysis," *Total Quality Management*, 1, 101–113.

—— (1992), "Split-plot Designs for Robust Product Experimentation," *Journal of Applied Statistics*, 19, 3–26.

—— (in press), "Designing Products That Are Robust to the Environment," *Total Quality Management*, 3.

Box, G. E. P., and Meyer, R. D. (1986), "Dispersion Effects From Fractional Designs," *Technometrics*, 28, 19–27.

Box, G. E. P., and Wilson, K. B. (1951), "On the Experimental Attainment of Optimum Conditions," *Journal of the Royal Statistical Society*, Ser. B., 13, 1–45.

Brayton, R. K., Hachtel, G. D., and Sangiovanni-Vincentelli, A. L. (1981), "A Survey of Optimization Techniques for Integrated-Circuit Design," *Proceedings of the IEEE*, 69, 1334–1362.

Buck, R., Liu, L., Nazaret, W., Sacks, J., and Welch, W. J. (1989), "A Sequential Statistical Strategy for Optimizing Circuit Designs: Multiple Performances," Technical Report, University of Illinois, Dept. of Statistics.

Carroll, R. J., and Ruppert, D. (1988), *Transformation and Weighting in Regression*, New York: Chapman and Hall.

Cox, D. R. (1958), *Planning of Experiments*, New York: John Wiley.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991), "Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments," *Journal of the American Statistical Association*, 86, 953–963.

Daniel, C. (1976), *Applications of Statistics to Industrial Experimentation*, New York: John Wiley.

Easterling, R. G. (1985), "Discussion of Kacker's Paper," *Journal of Quality Technology*, 17, 191–192.

Fieller, E. C. (1954), "Some Problems in Interval Estimation," *Journal of The Royal Statistical Society*, Ser. B, 16, 175–185.

Franklin, M. F. (1985), "Selecting Defining Contrasts and Confounded Effects in p^{n-m} Experiments," *Technometrics*, 27, 165–172.

Franklin, M. F., and Bailey, R. A. (1977), "Selection of Defining Contrasts and Confounded Effects in Two-Level Experiments," *Journal of the Royal Statistical Society*, Ser. C, 26, 321–26.

Freeny, A. E., and Nair, V. N. (in press), "Robust Parameter Design With Uncontrolled Noise Variables," *Statistica Sinica*, 2.

Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines" (with discussion), *The Annals of Statistics*, 19, 1–141.

Fries, A., and Hunter, W. G. (1980), "Minimum Aberration $2^{(k-p)}$ Designs," *Technometrics*, 22, 601–608.

Fuller, E. R., Jr., Krause, R. F., Kelly, J. F., Cranmer, D. C., Lagergren, E. S., Kacker, R. N., Barta, J., Jahn, P. F., and Scheyer, G. (1991), "Microstructure, Mechanical Properties and Machining Performance of Alumina-Silicon Carbide Whiskers Cutting Tools," technical report, National Institute of Standards and Technology, Ceramics Division, Gaithersburg, MD.

Fung, C. A. (1986), "Statistical Topics in Off-line Quality Control," unpublished Ph.D. thesis, University of Wisconsin-Madison, Dept. of Statistics.

Gosset, W. S. (1986), *Student's Collected Papers*, Cambridge, U.K.: Biometrika Trustees.

Greenfield, A. A. (1976), "Selection of Defining Contrasts in Two-Level Experiments," *Journal of the Royal Statistical Society*, Ser. C, 25, 64–67.

Grove, D. M., and Davis, T. P. (1991), "Taguchi's Idle Column Method," *Technometrics*, 33, 349–354.

Hamada, M. (in press), "An Explanation and Criticism of Minute Accumulating Analysis: Taguchi's Method for Life Test Data," *Journal of Quality Technology*, 24.

Hamada, M., and Tse, S. K. (in press), "On Estimability Problems in Industrial Experiments With Censored Data," *Statistica Sinica*, 2.

Hamada, M., and Wu, C. F. J. (1990), "A Critical Look at Accumulation Analysis and Related Methods" (with discussion), *Technometrics*, 32, 119–130.

———— (1991), "Analysis of Censored Data From Highly Fractionated Experiments," *Technometrics*, 33, 25–38.

———— (in press), "Analysis of Designed Experiments With Complex Aliasing," *Journal of Quality Technology*, 24.

Hellstrand, C. (1989), "The Necessity of Modern Quality Improvement and Some Experiences With its Implementation in the Manufacture of Rolling Bearings," *Proceedings of the Transactions of the Royal Society, London*, A 327, 529–537.

Hsieh, C. C., Oh, K. P., and Oh, H. L. (1990), "A Design Technique for Minimizing Net Build Errors," in *Proceedings of 1990 Winter Annual Meeting of the American Society of Mechanical Engineers*, Dallas, TX: American Society of Mechanical Engineers, pp. 61–69.

Iman, R. L., and Conover, W. J. (1982), "A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables," *Communications in Statistics, Part B—Simulation and Computation*, 11, 311–334.

Juran, J. M. (1979), *Quality Control Handbook* (3rd ed.), New York: McGraw-Hill.

Juran, J. M., and Gryna, F. M. (1980), *Quality Planning and Analysis*, New York: McGraw-Hill.

Kacker, R. N. (1985), "Off-Line Quality Control, Parameter Design, and the Taguchi Method," *Journal of Quality Technology*, 17, 176–209.

Kacker, R. N., and Shoemaker, A. C. (1986), "Robust Design: A Cost Effective Method for Improving Manufacturing Process," *AT&T Technical Journal*, 65, 39–50.

Kacker, R. N., and Tsui, K.-L. (1990), "Interaction Graphs: Graphical Aids for Planning Experiments," *Journal of Quality Technology*, 22, 1–14.

Kempthorne, O. (1952), *The Design and Analysis of Experiments*, Malabar, FL: Robert E. Krieger.

Kusaba, I. (1988), "Statistical Methods in Japanese Quality Control," *Societas Qualitatis*, 2, 1–3.

Leon, R. V., Shoemaker, A. C., and Kacker, R. N. (1987), "Performance Measure Independent of Adjustment: An Explanation and Extension of Taguchi's Signal to Noise Ratio," *Technometrics*, 29, 253–285.

Leon, R. V., and Wu, C. F. J. (in press), "A Theory of Performance Measures in Parameter Design," *Statistica Sinica*, 2.

Liggett, W. (1991), "Project Planning for Sequential Parameter Design Experiments," technical report, Statistical Engineering Division, NIST, Gaithersburg, MD.

Lin, K.-K., and Spanos, C. J. (1990), "Statistical Equipment Modeling for VLSI Manufacturing: An Application for LPCVD," *IEEE Transactions on Semiconductor Manufacturing*, 3, 216–229.

Lorenzen, T. J., and Truss, L. T. (1990), "Anatomy of DEXPERT—an Expert System for the Design of Experiments," Report GMR-7111, General Motors, Mathematics Dept.

Lorenzen, T. J., and Villalobos, M. A. (1990), "Understanding Robust Design, Loss Functions, and Signal to Noise Ratios," in *Proc. Sec. on Stat. Ed., ASA*, pp. 74–83.

Lucas, J. M. (1985), Comment on "Off Line Quality Control, Parameter Design, and the Taguchi Method," by R. N. Kacker, *Journal of Quality Technology*, 17, 195–197.

———— (1991), "Achieving a Robust Process Using Response Surface Methodology," technical report, E. I. DuPont de Nemours and Company, Newark, DE.

Lyon, G. E., Snelick, R. D., and Kacker, R. N. (1991), "Time-Perturbation Tuning of Multiple Instruction, Multiple Data Stream Supercomputer Programs," technical report, NIST Computer Science Laboratory, Gaithersburg, MD.

McCullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models* (2nd ed.), London: Chapman and Hall.

McKay, M. D., Conover, W. J., and Beckman, R. J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code," *Technometrics*, 21, 239–245.

Mandel, J., and Stiehler, R. D. (1954), "Sensitivity—A Criterion for the Comparison of Methods of Test," *Journal of Research of the National Bureau of Standards*, 53, 155–159.

Michaels, S. E. (1964), "The Usefulness of Experimental Design" (with discussion), 13, 221–235.

Miller, A., and Wu, C. F. J. (1991), "Improving a Calibration System Through Designed Experiments," Technical Report, University of Waterloo, IIQP.

Montgomery, D. C. (1991), "Using Fractional Factorial Designs for Robust Design Process Development," *Quality Engineering*, 3, 193–205.

Morrison, S. J. (1957), "The Study of Variability in Engineering Design, *Applied Statistics*, 6, 133–138.

Myers, R. H., and Carter, W. H., Jr. (1973), "Response Surface Techniques for Dual Response Systems," *Technometrics*, 15, 301–317.

Nair, V. N. (1986), "Testing in Industrial Experiments With Ordered Categorical Data" (with discussion), *Technometrics*, 28, 283–291.

Nair, V. N., and Pregibon, D. (1986), "A Data Analysis Strategy for Quality Engineering Experiments," *AT&T Technical Journal*, 73–84.

——— (1988), "Analyzing Dispersion Effects From Replicated Factorial Experiments," *Technometrics*, 30, 247–257.

Nair, V. N., and Shoemaker, A. C. (1990), "The Role of Experimentation in Quality Engineering: A Review of Taguchi's Contributions," in *Statistical Design and Analysis of Industrial Experiments*, ed. S. Ghosh, New York: Marcel Dekker, pp. 247–277.

Nazaret, W., and Liu, L. (1990), "Computer Aided Design for Quality (CADQ)," *AT&T Technical Journal*, 69, 46–60.

Nelder, J. A., and Lee, Y. (1991), "Generalized Linear Models for the Analysis of Taguchi-type Experiments," *Applied Stochastic Models and Data Analysis*, 7, 107–120.

Nelder, J. A., and Pregibon, D. (1987), "An Extended Quasi-likelihood Function," *Biometrika*, 74, 221–232.

Phadke, M. S. (1989), *Quality Engineering Using Robust Design*, Englewood Cliffs, NJ: Prentice-Hall.

Phadke, M. S., and Dehnad, K. (1988), "Optimization of Product and Process Design for Quality and Cost," *Quality and Reliability Engineering International*, 4, 105–112.

Phadke, M. S., Kacker, R. N., Speeney, D. V., and Frieco, J. J. (1983), "Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design," *The Bell System Technical Journal*, 1, 1273–1309.

Phadke, M. S., and Taguchi, G. (1987), "Selection of Quality Characteristics and S/N Ratios for Robust Design," in *Conference Record, GLOBECOM 87 Meeting, IEEE Communication Society* (Tokyo, Japan, November 1987), pp. 1002–1007.

Plackett, R. L., and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika*, 33, 305–325.

Quinlan, J. (1985), "Product Improvement by Application of the Taguchi Methods," in *Third Supplier Symposium on Taguchi Methods*, Dearborn, MI: American Supplier Institute, pp. 11–16.

Rosenblatt, A., and Watson, G. F. (1991), "Concurrent Engineering," *IEEE Spectrum*, July, 22–37.

Sacks, J., Schiller, S. B., and Welch, W. J. (1989), "Designs for Computer Experiments," *Technometrics*, 31, 41–47.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), "Design and Analysis of Computer Experiments" (with discussion), *Statistical Science*, 4, 409–435.

Shewhart, W. A. (1931), *Economic Control of Quality of Manufacturing Processes*, Milwaukee, WI: American Society for Quality Control.

Shoemaker, A. C., Tsui, K. L., and Wu, C. F. J. (1991), "Economical Experimentation Methods for Robust Design," *Technometrics*, 33, 415–427.

Taguchi, G. (1959), "Linear Graphs for Orthogonal Arrays and Their Application to Experimental Design, With the Aid of Various Techniques," *Reports of Statistical Application Research JUSE*, 6, 1–43.

——— (1960), "Table of Orthogonal Arrays and Linear Graphs," *Reports of Statistical Application Research, JUSE*, 6, 1–52.

——— (1986), *Introduction to Quality Engineering*, White Plains, NY: UNIPUB/Kraus International.

——— (1987), *System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*, White Plains, NY: UNIPUB/Kraus International.

Taguchi, G., and Phadke, M. S. (1984), "Quality Engineering Through Design Optimization," in *Conference Record, GLOBECOM 84 Meeting, IEEE Communications Society* (Atlanta, GA, November), pp. 1106–1113.

Taguchi, G., and Wu, Y. (1985), *Introduction to Off-Line Quality Control*, Nagoya, Japan: Central Japan Quality Control Association.

Tsui, K.-L. (1988), "Strategies for Planning Experiments Using Orthogonal Arrays and Confounding Tables," *Quality and Reliability Engineering International*, 4, 113–122.

——— (1989), "An Automatic Experiment Planner for Robust Design," technical report, AT&T Bell Laboratories.

Vining, G. G., and Myers, R. H. (1990), "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach, *Journal of Quality Technology*, 22, 38–45.

Wang, J. C., and Wu, C. F. J. (1991), "An Approach to the Construction of Asymmetrical Orthogonal Arrays," *Journal of the American Statistical Association*, 86, 450–456.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992), "Screening, Predicting, and Computer Experiments," *Technometrics*, 34, 15–25.

Welch, W. J., and Sacks, J. (1991), "A System for Quality Improvement Via Computer Experiments," *Communications in Statistics—Theory and Methods*, 20, 477–495.

Welch, W. J., Yu, T. K., Kang, S. M., and Sacks, J. (1990), "Computer Experiments for Quality Control by Parameter Design," *Journal of Quality Technology*, 22, 15–22.

Wernimont, G. (1977), "Ruggedness Evaluation of Test Procedures," *Standardization News*, 5, 13–16.

Wu, C. F. J., and Chen, Y. Y. (1992), "A Graph-aided Method for Planning Two-level Experiments When Certain Interactions Are Important," *Technometrics*, 34, 162–175.

Wu, C. F. J., Mao, S. S., and Ma, F. S. (1990), "SEL: A Search Method Based on Orthogonal Arrays," in *Statistical Design and Analysis of Industrial Experiments*, ed. S. Ghosh, New York: Marcel Dekker, pp. 279–310.

Yano, H. (1991), *Metrological Control: Industrial Measurement Management*, Tokyo: Asian Productivity Organization.

Youden, W. J. (1961a), "Experimental Design and ASTM committee," *Materials Research and Standards*, 1, 862–867. Reprinted in *Precision Measurement and Calibration* (Vol. 1, Special Publication 300), Gaithersburg, MD: National Bureau of Standards, 1969, ed. H. H. Ku.

——— (1961b), "Physical Measurement and Experimental Design," *Collogues Internationaux de Centre national de la Recherche Scientifique No. 110, le Plan d'Experiences*, 115–128. Reprinted in *Precision Measurement and Calibration* (Vol. 1, Special Publication 300), Gaithersburg, MD: National Bureau of Standards, 1969, ed. H. H. Ku.

Yu, T. K., Kang, S. M., Sacks, J., and Welch, W. J. (in press), "Parametric Yield Optimization of CMOS Analog Circuits by Quadratic Statistical Circuit Performance Models," *International Journal of Circuit Theory and Applications*, 20.