# Bayesian Optimal Single Arrays for Robust Parameter Design

Lulu Kang and V. Roshan Joseph
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA30332-0205, USA
lkang@isye.gatech.edu & roshan@isye.gatech.edu

**Abstract**

It is critical to estimate control-by-noise interactions in robust parameter design. This can be achieved by using a cross array, which is a cross product of a design for control factors and another design for noise factors. However, the total run size of such arrays can be prohibitively large. To reduce the run size, single arrays are proposed in the literature, where a modified effect hierarchy principle is used for the optimal selection of the arrays. In this article, we argue that effect hierarchy is a property of the system and cannot be altered depending on the objective of an experiment. We propose a Bayesian approach to develop single arrays which incorporate the importance of control-by-noise interactions without altering the effect hierarchy. The approach is very general and places no restrictions on the number of runs or levels or type of factors or type of designs. A modified exchange algorithm is proposed for finding the optimal single arrays. We also explain how to design experiments with internal noise factors. The advantages of the proposed approach are illustrated using several examples.

KEY WORDS: Design of experiment; Exchange algorithms; Internal noise; Quality improvement; Variation reduction.

April 10, 2008

# 1 INTRODUCTION

Robust parameter design is a useful technique for quality improvement. The main idea of this technique is to divide the factors in the system into two groups: control factors and noise factors. Then, the settings of the control factors are chosen so that the sensitivity of the response to the noise factors is minimized. This in turn minimizes the transmitted variance to the response from the noise factors and thus, improving the quality. Clearly, robust parameter design can be successful only if the control factors are interacting with the noise factors. Therefore, the experimental design for robustness studies should be capable of estimating the control-by-noise interactions. This aspect makes such designs different from the traditional designs.

Taguchi (1987) proposed to use *cross array*, which is a cross product of a design for control factors (control array) and another design for noise factors (noise array). Thus, in cross arrays, each run in the noise array is repeated for each run in the control array. The advantage of cross arrays is that the interactions between control factors and noise factors can be estimated (see Wu and Hamada 2000, Section 10.7). The disadvantage is that because of the crossing of two arrays, the total run size of the design is very large.

To overcome the drawback of large run size of cross arrays, Welch et al. (1990) and Shoemaker, Tsui, and Wu (1991) proposed to use single arrays. *Single arrays* are fractional factorial designs that incorporate both control factors and noise factors. Single array is a more general concept in the sense that any cross array can be converted to a single array, but in general, the converse is not true. Designing single arrays is more difficult than a traditional fractional factorial design because of the special importance given to control-by-noise interactions.

The traditional fractional factorial designs are based on the fundamental principle of effect hierarchy. The *effect hierarchy principle* states that lower order effects are more likely to be important than higher order effects and effects of the same order are equally likely to be important (e.g., Wu and Hamada 2000, Section 3.5). To explain this principle in the context of robust parameter design, let us denote a control factor by $C$ and a noise factor by $n$. Then, according to the hierarchy principle

$$\{C, n\} \succ \{CC, Cn, nn\} \succ \{CCC, CCn, Cnn, nnn\}, \ldots,$$

where $\succ$ denotes "more important than". This does not seem to agree with the objective of robust parameter design, because control-by-noise interactions are considered to be more important than other types of interactions. A quick remedy for this problem seems to be to modify the hierarchy principle so as to satisfy the objectives of robust parameter design. Wu and Zhu (2003) proposed the following modified ordering:

$$\{C, n, Cn\} \succ \{CC, CCn\} \succ \{CCnn, Cnn, nn\}, \ldots,$$

whereas Bingham and Sitter (2003) proposed the ordering

$$\{C, n\} \succ \{Cn\} \succ \{CC, nn\} \succ \{CCn, Cnn\}, \ldots.$$

See also Zhu, Zeng, and Jennings (2007). Although this approach produces reasonably good single arrays, there is a limitation.

Effect hierarchy is a property of an engineering system and therefore, cannot be changed depending on the objective of an experiment. For example, consider the leaf spring experiment given in Wu and Hamada (2000). Four control factors (high heat temperature, heating time, transfer time, and hold down time) and one noise factor (quench oil temperature) are studied in this experiment. The interactions among these factors are completely dependent on the underlying physics of the heat treatment process. In general, these interactions tend to follow the effect hierarchy principle. Although the objective is to achieve robustness, we cannot expect that control-by-noise interactions are going to be "more significant" than the other interactions. Of course, they are "more important" in terms of achieving the objectives of the experiment, but there is no way we can alter the already fixed effect hierarchy of the heat treatment process. In this article, we propose a methodology that can incorporate the importance of control-by-noise interactions without altering the effect hierarchy principle.

A nice feature of our approach is that there is no restriction on the number of levels of the factors or the number of runs in the experiment. On the other hand, Bingham and Sitter (2003) and Wu and Zhu (2003) focus only on $2^{p-k}$ fractional factorial experiments. The response surface designs in Borror, Montgomery, and Myers (2002), Ginsburg and Ben-Gal (2006), and Del Castillo et al. (2007) are general and can entertain different number of factor levels and runs. However, such designs are large and do not incorporate effect hierarchy principle.

The paper is organized as follows. In Section 2, we propose a Bayesian approach to design single arrays that incorporates the robustness objectives while honoring the effect hierarchy principle. Application of the methodology to two-level experiments is discussed in Section 3. In this section, we also develop an exchange algorithm that can be used for finding the optimal design. Some examples of optimal single arrays are then provided to compare with the existing designs. In Section 4, we generalize our method to mixed-level experiments. In Section 5, we discuss how to design experiments with internal noise factors; a topic that has received scant attention in the literature. We finally conclude our paper with some remarks in Section 6. All proofs are provided in the Appendix.

# 2 OPTIMAL DESIGN CRITERION

Our objective is to develop a single array that recognizes the importance of various effects while maintaining effect hierarchy. A Bayesian formulation is suitable for this purpose. We put a prior on the effects so as to reflect effect hierarchy and develop an optimal design criterion that gives more importance to the effects of interest.

Let there be $k_C$ control factors $\boldsymbol{x} = (x_1, x_2, \ldots, x_{k_C})$ and $k_n$ noise factors $\boldsymbol{z} = (z_1, z_2, \ldots, z_{k_n})$. The response $y$ is related to the control and noise factors through

$$y = f(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}) + \epsilon, \tag{1}$$

where $\epsilon \sim N(0, \sigma^2)$ is the random error caused by the unobserved noise factors and $\boldsymbol{\beta}$ is a set of unknown parameters in the model. Although the noise factors are fixed during experimentation, in the actual process they are random. Let $E(z_i) = 0$ and $var(z_i) = \sigma_z^2$ for $i = 1, \ldots, k_n$. By choosing the same variance for all the noise factors, we implicitly assume that the noise factor levels for the experiment are chosen corresponding to the same quantiles of their respective distributions. We also assume them to be independent. When the specific form of the transfer function $f$ is not known, it is convenient to use a linear model: $f(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}) = \sum_i \beta_i u_i(\boldsymbol{x}, \boldsymbol{z})$, where $u_i$'s are known functions and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots)$ are unknown parameters. Let the prior distribution for $\boldsymbol{\beta}$ be $N(\boldsymbol{\mu}, \tau^2 \boldsymbol{R})$. We now choose $\boldsymbol{\mu}$ and $\boldsymbol{R}$ in such a way that the effect hierarchy is maintained. This can be easily done using the results in Joseph (2006) and Joseph and Delaney (2007).

Now that effect hierarchy is already incorporated into the model, we only need to focus on deriving an optimal design criterion that satisfy the objectives of robust parameter design. First, approximate the transfer function using a first order Taylor series expansion:

$$f(\boldsymbol{x}, z, \boldsymbol{\beta}) \approx f(\boldsymbol{x}, \mathbf{0}, \boldsymbol{\beta}) + \nabla_z f(\boldsymbol{x}, \mathbf{0}, \boldsymbol{\beta})' z. \tag{2}$$

We assume relatively small $\sigma_z$ to make the first order expansion a reasonable approximation. Then, the mean and variance of the response are given by

$$E(y) \approx f(\boldsymbol{x}, \mathbf{0}, \boldsymbol{\beta}), \tag{3}$$

$$var(y) \approx \sigma_z^2 \nabla_z f(\boldsymbol{x}, \mathbf{0}, \boldsymbol{\beta})' \nabla_z f(\boldsymbol{x}, \mathbf{0}, \boldsymbol{\beta}) + \sigma^2. \tag{4}$$

For a nominal-the-best characteristic, the objective is to achieve the mean at target with minimum variation. Adjusting the mean to target is an easy job, but reducing the variance is very challenging. Therefore, estimation of the mean is not as important as estimation of variance in robust parameter design. In fact, in most cases adjustment factors can be found from engineering knowledge and can be used for adjusting the mean to any desired level (see Joseph 2007). Thus, we propose to design the experiment to efficiently estimate the variance.

As can be seen in (4), the variance can be efficiently estimated if we can get an efficient estimate of the gradient $\nabla_z f(\boldsymbol{x}, \mathbf{0}, \boldsymbol{\beta})$. Let $\widehat{\boldsymbol{\beta}}$ be the Bayes estimate (posterior mean) of $\boldsymbol{\beta}$. Then, the difference between the true gradient and the estimated gradient is

$$d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}) = \nabla_z f(\mathbf{x}, \mathbf{0}, \boldsymbol{\beta}) - \nabla_z f(\mathbf{x}, \mathbf{0}, \widehat{\boldsymbol{\beta}}), \tag{5}$$

We want this difference to be as small as possible. Therefore, we should minimize

$$\begin{aligned} l(\boldsymbol{x}) &= E\{d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}})' d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}})\} \\ &= \int_{\boldsymbol{y}} \int_{\boldsymbol{\beta}} d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}})' d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}) p(\boldsymbol{\beta} | \mathbf{y}) d\boldsymbol{\beta} p(\mathbf{y}) d\mathbf{y}. \end{aligned} \tag{6}$$

The data $\boldsymbol{y}$ depends on the experimental design $D$. Therefore, we can choose a $D$ such that $l(\boldsymbol{x})$ is a minimum. However, $l(\boldsymbol{x})$ varies with $\boldsymbol{x}$. We want $l(\boldsymbol{x})$ to be small over the entire experimental region for control factors $\mathscr{X}$, which we consider as a discrete space containing all the candidates points for the experimental design. Therefore, we may choose $D$ that minimizes the average loss over $\mathscr{X}$. Thus, our optimal design criterion is to find $D$ to minimize

$$L(D) = \frac{1}{|\mathscr{X}|} \sum_{\boldsymbol{x} \in \mathscr{X}} l(\boldsymbol{x}), \tag{7}$$

where $|\mathscr{X}|$ is the number of points in the set $\mathscr{X}$. In the following sections, we apply this general method to two- and mixed-level experiments.

# 3 TWO-LEVEL EXPERIMENTS

## 3.1 Optimal Design Criterion

Consider the case of two-level factors. Let the two levels be $-1$ and $1$. The full linear model can be written as

$$\begin{aligned} f(\boldsymbol{x}, z, \boldsymbol{\beta}) = \mu_0 + &\left\{ \beta_0^0 + \sum_{i=1}^{k_C} \beta_i^0 x_i + \ldots + \beta_{12\ldots k_C}^0 x_1 x_2 \ldots x_{k_C} \right\} + \\ &+ \sum_{j=1}^{k_n} \left\{ \beta_0^j + \sum_{i=1}^{k_C} \beta_i^j x_i + \ldots \right\} z_j + \ldots + \left\{ \beta_0^{2_n^k} + \sum_{i=1}^{k_C} \beta_i^{2_n^k} x_i + \ldots \right\} z_1 z_2 \ldots z_{k_n}. \end{aligned} \tag{8}$$

Here $\mu_0$ is a constant which is introduced only for simplifying the prior distribution. Let m.e. denote main effects ($\beta_i^0$ for $i = 1, \ldots, k_C$ and $\beta_0^j$ for $j = 1, \ldots, k_n$), 2fi denote two-factor interactions, etc. We use the following prior for $\boldsymbol{\beta}$ proposed in Joseph (2006):

$$\begin{aligned} \beta_0^0 &\sim N(0, \tau^2), \\ \beta_{m.e.} &\sim N(0, \tau^2 r), \\ \beta_{2fi} &\sim N(0, \tau^2 r^2), \\ &\vdots \end{aligned} \tag{9}$$

and they are all independent, where $r$ is a value between 0 and 1. Thus, $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \tau^2 \boldsymbol{R})$, where $\boldsymbol{R}$ is a diagonal matrix with entries $1, r, \ldots, r, r^2, \ldots, r^{k_C + k_n}$. In this prior, the variances of effects decrease geometrically as the order of the effects increase. Therefore, the probability that an effect becomes significant decreases as the order increases, justifying effect hierarchy.

Consider an experiment with $m$ runs. Let $\boldsymbol{D}$ be the $m \times (k_C + k_n)$ design matrix and $\boldsymbol{U}_D$ be the $m \times 2^{k_C + k_n}$ model matrix corresponding to the linear model in (8). Let $\boldsymbol{y} = (y_1, \ldots, y_m)$ be the data from the experiment. Then, the posterior variance of $\boldsymbol{\beta}$ is given by

$$var(\boldsymbol{\beta}|\boldsymbol{y}) = \tau^2\boldsymbol{R} - \tau^2\boldsymbol{R}\boldsymbol{U}_D'\left(\boldsymbol{U}_D\boldsymbol{R}\boldsymbol{U}_D' + \frac{\sigma^2}{\tau^2}\boldsymbol{I}_n\right)^{-1}\boldsymbol{U}_D\boldsymbol{R}. \tag{10}$$

The objective function in (7) can be easily computed using the following result.

**Proposition 1.** *The objective function defined in* (7) *is*

$$L(\boldsymbol{D}) = tr(\boldsymbol{A}var(\boldsymbol{\beta}|\boldsymbol{y})), \tag{11}$$

*where* $\boldsymbol{A}$ *is a diagonal matrix, whose diagonal entries are 1 if they correspond to effects involving one and only one noise factor, i.e.,* $n, Cn, CCn, CCCn, \ldots$*, and 0 otherwise.*

To understand the foregoing objective function better, consider a small example with two control factors and two noise factors. The model is

$$\begin{aligned} y = \mu_0 &+ \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 z_1 + \beta_4 z_2 + \beta_5 x_1 z_1 + \beta_6 x_1 z_2 + \beta_7 x_2 z_1 + \beta_8 x_2 z_2 \\ &+ \beta_9 x_1 x_2 + \beta_{10} z_1 z_2 + \beta_{11} x_1 x_2 z_1 + \beta_{12} x_1 x_2 z_2 + \beta_{13} x_1 z_1 z_2 + \beta_{14} x_2 z_1 z_2 + \beta_{15} x_1 x_2 z_1 z_2 + \epsilon, \end{aligned} \tag{12}$$

Denote $V_i = var(\beta_i|\boldsymbol{y})$. Using (11), we obtain

$$L(\boldsymbol{D}) = V_3 + V_4 + V_5 + V_6 + V_7 + V_8 + V_{11} + V_{12}.$$

We can see that the objective function includes only the noise main effects ($\beta_3$ and $\beta_4$) and interactions between control and noise factors ($\beta_5, \beta_6, \beta_7, \beta_8, \beta_{11}$, and $\beta_{12}$). The control main effects ($\beta_1$ and $\beta_2$) and control-by-control interaction ($\beta_9$) affect only the mean and therefore, are not needed for minimizing variance. The higher order noise effects ($\beta_{13}, \beta_{14}$, and $\beta_{15}$) are neglected in computing the variance. Thus, the optimal design criterion includes only those effects that are important for achieving robustness. Note that effect hierarchy is already incorporated in the prior information and therefore, the foregoing Bayesian method is able to separately incorporate both "importance" and "significance" of effects, which is not the case with the existing frequentist methods.

Because the loss function $L(\boldsymbol{D})$ does not have any interpretation, it is more convenient to transform it to the following utility function

$$U(\boldsymbol{D}) = \frac{tr\,(\boldsymbol{A}var(\boldsymbol{\beta})) - tr\,(\boldsymbol{A}var(\boldsymbol{\beta}|\mathbf{y}))}{tr\,(\boldsymbol{A}var(\boldsymbol{\beta}))}. \tag{13}$$

Thus, if the design is completely noninformative ($m = 0$), then $U(\boldsymbol{D}) = 0$ and if it is completely informative ($m = 2^{k_C+k_n}$), then $U(\boldsymbol{D}) = 1$. Therefore, the value of $U(\boldsymbol{D})$ clearly shows how good the design is. We obtain

$$U(\boldsymbol{D}) = \frac{tr\left(\boldsymbol{ARU}_D'\left(\boldsymbol{U}_D\boldsymbol{RU}_D' + \sigma^2/\tau^2\boldsymbol{I}_n\right)^{-1}\boldsymbol{U}_D\boldsymbol{R}\right)}{tr(\boldsymbol{AR})}. \tag{14}$$

The optimal design can be found by maximizing $U(\boldsymbol{D})$.

There are several parameters ($\sigma^2, \tau^2$, and $r$) that need to be specified before we can find the optimal design. The problem is, in most cases, we do not know the "best" values of these parameters before conducting the experiment. Therefore, some reasonable and meaningful values should be chosen for designing the experiment. As argued by Joseph (2006), by assuming a high signal-to-noise ratio we can neglect the ratio $\sigma^2/\tau^2$. Then, the criterion reduces to maximizing $tr(\boldsymbol{ARU}_D'(\boldsymbol{U}_D\boldsymbol{RU}_D')^{-1}\boldsymbol{U}_D\boldsymbol{R})$. Note that we omitted the denominator in (14), because it does not depend on $\boldsymbol{D}$. Now we only need to specify the value of $r$. Li, Sudarsanam, and Frey (2006) did a meta-analysis of 113 data sets from published experiments. They found that the median strength of two-factor interactions is 1/4th of the median strength of main effects and the median strength of three-factor interactions is half of the median strength of two-factor interactions. Based on their finding, we think it is reasonable to choose $r = 1/3$ in the absence of any other prior knowledge about the particular product/process under investigation.

Having specified all the unknown parameters in the objective function, we can proceed to find the optimal design for a given number of runs $m$. The Bayesian approach is capable of producing an optimal design for any $m$. This is definitely an advantage over the frequentist methods. However, we should realize that a Bayesian approach cannot do any magic. If the run size is too small, then the results heavily rely on the prior information as there is little information from the experiments. Therefore, even though we use a Bayesian approach, it is important to choose a reasonable number of runs for the experiment. We recommend that the design should be capable of estimating at least the grand mean, control main effects, noise main effects, and two-factor control-by-noise interactions in the frequentist sense. A necessary condition for this is

$$m \geq 1 + k_C + k_n + k_C k_n = (1 + k_n)(1 + k_C). \tag{15}$$

Note that although our objective function includes only the $n$ and $Cn$ effects, we additionally included $I$ and $C$ effects when calculating the minimum number of runs. This is because, if the grand mean or the control main effects are not estimable, i.e., $I = C$ or $C_1 = C_2$, then $n = Cn$ or $C_1 n = C_2 n$, which should not happen.

## 3.2 Exchange Algorithm

Exchange algorithms are the most common form of computer design search algorithms. See Nguyen and Miller (1992) for a review of some early versions of exchange algorithm for constructing D-optimal designs. The basic idea of exchange algorithm is to randomly choose an initial design and perform row-wise exchanges of some points from a candidate set of feasible design points until the objective function is optimized. There have been many new types of exchange algorithms developed for some other optimal design criteria, such as columnwise-pairwise exchange algorithm (Li and Wu, 1997) and coordinate-exchange algorithm (Meyer and Nachtsheim, 1995).

Although these algorithms are very effective in optimizing most frequentist based optimal design criteria, it is hard to implement them with our Bayesian optimal design criterion. Therefore, we propose a modified exchange algorithm that takes advantage of the special matrix form in our objective function. Moreover, instead of randomly choosing an initial design, we construct our initial design in a "greedy" fashion to accelerate the convergence of the later "exchange" part of the algorithm. In the simulations we carried out, this modified algorithm is found to perform quite well.

For this section, $\boldsymbol{x}$ is used for generically denoting a point in the design space containing *both* control and noise factors. Let $D = \{\mathbf{x}_1, \ldots, \boldsymbol{x}_l\}$ be the set of points that are currently selected as design points, where $l \leq m$. The remaining $2^{k_C+k_n} - l$ points of the design space form the candidate set. The model matrix generated for the points in $D$ is denoted as $\boldsymbol{U}_D$. If a new point $\boldsymbol{x}$ is added into $D$, then it introduces a new row in the model matrix $\boldsymbol{U}_D$. Denote the new row in $\boldsymbol{U}_D$ by $\boldsymbol{F}' = (1, u_1, u_2, \ldots)'$ which consists of all the main effects and interactions. The new model matrix is $\tilde{\boldsymbol{U}}_D = (\boldsymbol{U}_D', \boldsymbol{F})'$.

Let $\boldsymbol{M} = \boldsymbol{R}\boldsymbol{U}_D'(\boldsymbol{U}_D\boldsymbol{R}\boldsymbol{U}_D')^{-1}\boldsymbol{U}_D\boldsymbol{R}$, and $\boldsymbol{M}_x = \boldsymbol{R}\tilde{\boldsymbol{U}}_D'(\tilde{\boldsymbol{U}}_D\boldsymbol{R}\tilde{\boldsymbol{U}}_D')^{-1}\tilde{\boldsymbol{U}}_D\boldsymbol{R}$. As shown in the Appendix, we can update $\boldsymbol{M}_x$ in the following way:

$$\boldsymbol{M}_x = \boldsymbol{M} + \frac{1}{d}\left((\boldsymbol{R}-\boldsymbol{M})\boldsymbol{F}\right)\left((\boldsymbol{R}-\boldsymbol{M})\boldsymbol{F}\right)', \tag{16}$$

where $d = \boldsymbol{F}'(\boldsymbol{R}-\boldsymbol{M})\boldsymbol{F}$. Thus the objective function $tr(\boldsymbol{A}\boldsymbol{M})$ is increased to $tr(\boldsymbol{A}\boldsymbol{M}_x)$ by $\Delta(\boldsymbol{x}, D)$, where

$$\begin{aligned}\Delta(\boldsymbol{x}, D) &= \frac{1}{d}((\boldsymbol{R}-\boldsymbol{M})\boldsymbol{F})'\boldsymbol{A}((\boldsymbol{R}-\boldsymbol{M})\boldsymbol{F}) \\ &= \frac{1}{d}\sum_{i:A_{ii}=1}((\boldsymbol{R}-\boldsymbol{M})\boldsymbol{F})_i^2 .\end{aligned} \tag{17}$$

We do not need to worry about the numerical stability of the algorithm due to two observations: (i) $\boldsymbol{R}-\boldsymbol{M}$ is a positive definite matrix so that $d > 0$ and (ii) as long as no points in $D$ are replicated, $\boldsymbol{U}_D\boldsymbol{R}\boldsymbol{U}_D'$ is invertible. Because of (ii), we only search for designs where none of the rows are replicated. However, replication is useful for estimating the unknown variance $\sigma^2$. Therefore, if we can afford replications, then we simply concatenate the optimal unreplicated design as many times as the number of replicates. The algorithm is given below.

*step 0.* Generate the complete design space which contains $2^{k_C+k_n}$ points, the correlation matrix $\boldsymbol{R}$, and specify the matrix $\boldsymbol{A}$.

*step 1.* Randomly select $m_0$ points into design point set $D$. Update $\boldsymbol{U}_D$, $\boldsymbol{M}$, and objective value is $tr(\boldsymbol{A}\boldsymbol{M})$.

*step 2.* Initial design: evaluate the increment $\Delta(\boldsymbol{x}, D)$ in (17) for every point in the current candidate set. Add the candidate point that gives the largest $\Delta(\boldsymbol{x}, D)$ into $D$. Then update $D$ and the candidate set. There is no need to update $\boldsymbol{U}_D$. Instead, we can directly update $\boldsymbol{M}$ to $\boldsymbol{M}_x$ by using (16). Repeat this step until $m$ points have been selected into $D$.

*step 3.* Exchange part: for each point $\boldsymbol{x}_i$ in $D$, denote $D_{-i} = D\backslash\boldsymbol{x}_i$. Compute

$$\boldsymbol{M}_{-i} = \boldsymbol{R}\boldsymbol{U}_{D_{-i}}'(\boldsymbol{U}_{D-i}\boldsymbol{R}\boldsymbol{U}_{D_{-i}}')^{-1}\boldsymbol{U}_{D_{-i}}\boldsymbol{R}.$$

Evaluate the $\Delta(\boldsymbol{x}, D_{-i})$ for every candidate point $\boldsymbol{x}$. Choose the $\boldsymbol{x}^*$ that has $\Delta(\boldsymbol{x}^*, D_{-i}) > \Delta(\boldsymbol{x}_i, D_{-i})$ and also $\Delta(\boldsymbol{x}^*, D_{-i}) \geq \Delta(\boldsymbol{x}, D_{-i})$ for any $\boldsymbol{x}$ in the candidate set. Exchange $\boldsymbol{x}^*$ with $\boldsymbol{x}_i$. Then update $D$ and the candidate set. Update the objective function value to $tr(\boldsymbol{A}\boldsymbol{M}_{-i}) + \Delta(\boldsymbol{x}^*, D_{-i})$.

*step 4.* Repeat *step 3* until the objective value has been stabilized.

Like all exchange algorithms, the optimal design returned can be a local optimum. So we begin with a randomly selected $m_0$ points. If $m_0$ is too small, little randomness is introduced and it is hard to escape from a local optimum; if $m_0$ is close to $m$, then Step 2 plays such a small role that the exchange part converges very slowly. Our simulation study shows that $m_0 \in \{m/4, m/3, m/2\}$ are good choices. Many trials of the algorithm should be employed and the best design should be returned. When the number of factors is large, the storage of the model matrix $\boldsymbol{U}$ can be an issue. In such cases, the size of $\boldsymbol{U}$ can be reduced by using only up to two- or three-factor interactions. The code for generating the optimal design is available from the authors upon request.

## 3.3 Examples

*Example 1*: Miller et al. (1993) studied an experiment on the geometric distortion of drive gears in a heat treatment process. There are five control factors: carbon potential ($A$), operating model ($B$), last zone temperature ($C$), quench oil temperature ($D$), and quench oil agitation ($E$) and three noise factors: furnace track ($a$), tooth size ($b$), and part positioning ($c$) (denoted as $F$, $G$, and $H$ in the original paper). The original experiment used a cross array design with $16 \times 8 = 128$ runs. Miller et al. (1993) reanalyzed the data using single arrays of sizes 64 and 32 and obtained essentially the same results as the original experiment. Here we consider a much smaller single array. From (15), $m \geq 24$. Therefore, we choose $m = 24$. With $r = 1/3$, we obtain the optimal design that maximizes $U(\boldsymbol{D})$. The optimal design ($D_1$) is shown in Table 1.

Consider an alternative design $D_2$, which is a $D$-optimal design for the model containing only control and noise main effects and two-factor control-by-noise interactions (we created this using JMP 7.0). To compare the two designs from a frequentist view point, consider the average absolute correlation of an effect with the other effects (main effects and two-factor interactions). Table 2 shows the average of the average absolute correlations of effects within each group. Clearly the effects of the type $n$ and $Cn$ are less correlated with the other effects in $D_1$ than $D_2$ and therefore, $D_1$ is a better design for estimating the noise main effects and

Table 1: The optimal design $D_1$ and $D$-optimal design $D_2$ for Example 1

| | Optimal Design ($D_1$) | | | | | | | | $D$-Optimal Design ($D_2$) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| run | $A$ | $B$ | $C$ | $D$ | $E$ | $a$ | $b$ | $c$ | $A$ | $B$ | $C$ | $D$ | $E$ | $a$ | $b$ | $c$ |
| 1 | −1 | 1 | −1 | −1 | −1 | −1 | −1 | −1 | −1 | −1 | −1 | 1 | −1 | −1 | −1 | 1 |
| 2 | −1 | −1 | 1 | 1 | −1 | −1 | −1 | −1 | −1 | −1 | 1 | −1 | 1 | −1 | 1 | −1 |
| 3 | 1 | 1 | 1 | −1 | 1 | −1 | −1 | −1 | −1 | −1 | −1 | −1 | −1 | 1 | 1 | 1 |
| 4 | 1 | −1 | −1 | 1 | −1 | 1 | −1 | −1 | −1 | −1 | −1 | 1 | 1 | 1 | 1 | 1 |
| 5 | −1 | −1 | 1 | −1 | 1 | 1 | −1 | −1 | −1 | 1 | 1 | −1 | −1 | −1 | 1 | −1 |
| 6 | −1 | 1 | −1 | 1 | 1 | 1 | −1 | −1 | 1 | 1 | 1 | −1 | 1 | 1 | −1 | −1 |
| 7 | 1 | −1 | −1 | 1 | −1 | −1 | 1 | −1 | 1 | −1 | −1 | 1 | −1 | 1 | −1 | −1 |
| 8 | −1 | −1 | 1 | −1 | 1 | −1 | 1 | −1 | −1 | −1 | −1 | −1 | 1 | 1 | −1 | −1 |
| 9 | −1 | 1 | −1 | 1 | 1 | −1 | 1 | −1 | 1 | −1 | 1 | −1 | 1 | −1 | −1 | 1 |
| 10 | −1 | 1 | −1 | −1 | −1 | 1 | 1 | −1 | −1 | −1 | 1 | 1 | −1 | 1 | −1 | −1 |
| 11 | −1 | −1 | 1 | 1 | −1 | 1 | 1 | −1 | −1 | 1 | 1 | 1 | 1 | −1 | 1 | −1 |
| 12 | 1 | 1 | 1 | −1 | 1 | 1 | 1 | −1 | −1 | 1 | −1 | −1 | −1 | 1 | −1 | −1 |
| 13 | 1 | −1 | −1 | 1 | −1 | −1 | −1 | 1 | 1 | 1 | −1 | −1 | 1 | −1 | 1 | −1 |
| 14 | −1 | −1 | 1 | −1 | 1 | −1 | −1 | 1 | −1 | 1 | 1 | 1 | −1 | 1 | 1 | 1 |
| 15 | −1 | 1 | −1 | 1 | 1 | −1 | −1 | 1 | −1 | −1 | 1 | −1 | 1 | 1 | 1 | 1 |
| 16 | −1 | 1 | −1 | −1 | −1 | 1 | −1 | 1 | 1 | 1 | −1 | 1 | 1 | −1 | −1 | 1 |
| 17 | −1 | −1 | 1 | 1 | −1 | 1 | −1 | 1 | −1 | 1 | 1 | 1 | 1 | −1 | −1 | 1 |
| 18 | 1 | 1 | 1 | −1 | 1 | 1 | −1 | 1 | −1 | 1 | −1 | −1 | −1 | −1 | −1 | 1 |
| 19 | −1 | 1 | −1 | −1 | −1 | −1 | 1 | 1 | 1 | −1 | 1 | 1 | −1 | −1 | 1 | −1 |
| 20 | −1 | −1 | 1 | 1 | −1 | −1 | 1 | 1 | −1 | −1 | −1 | 1 | −1 | −1 | 1 | −1 |
| 21 | 1 | 1 | 1 | −1 | 1 | −1 | 1 | 1 | −1 | 1 | −1 | 1 | 1 | 1 | −1 | −1 |
| 22 | 1 | −1 | −1 | 1 | −1 | 1 | 1 | 1 | 1 | −1 | 1 | 1 | −1 | 1 | 1 | 1 |
| 23 | −1 | −1 | 1 | −1 | 1 | 1 | 1 | 1 | 1 | 1 | −1 | −1 | 1 | 1 | 1 | 1 |
| 24 | −1 | 1 | −1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | −1 | −1 | −1 | 1 |

control-by-noise interactions.

Figure 1 shows the relative efficiency $U(D_2)/U(D_1)$ for different values of $r$. We can see that $D_1$ is uniformly better than $D_2$ for all values of $r \in (0,1)$. Figure 2 shows the relative efficiency for different values of $\sigma^2/\tau^2 \in [0,5]$ with $r = 1/3$. $D_1$ is still uniformly better than $D_2$, but the improvement diminishes as $\sigma^2/\tau^2$ increases. We also studied the effect of number of runs. The utility function of the optimal design $U(D^*)$ for $m = 1,2,\ldots,256$ with $r = 1/3$ and $\sigma^2/\tau^2 = 0$ is shown in Figure 3. It can be seen that the value of $U(D^*)$ increases with $m$ and reaches 1 when the design is full factorial.

*Example 2*: Suppose an experiment involves five control factors $A, B, C, D, E$ and one noise factor $a$. From (15), $m \geq 12$. This time let us choose $m = 16$. Interestingly, for all $r \in (0,1)$, the

Table 2: Average absolute correlations of effects for $D_1$ and $D_2$ in Example 1

| Design | $C$ | $n$ | $Cn$ | $CC$ | $nn$ |
|---|---|---|---|---|---|
| Optimal design ($D_1$) | 0.1674 | 0.0152 | 0.0515 | 0.1722 | 0.0606 |
| D-Optimal design ($D_2$) | 0.1589 | 0.1496 | 0.1540 | 0.2137 | 0.1496 |

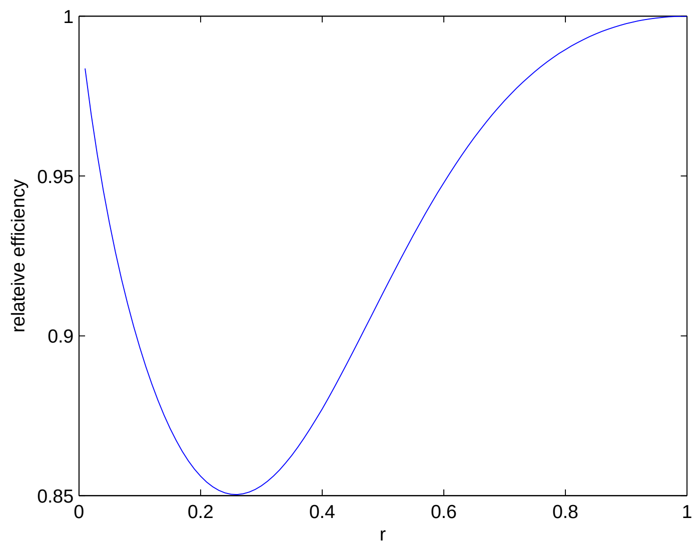


Figure 1: Relative Efficiency of $D_2$ to $D_1$ ($\sigma^2/\tau^2 = 0$).

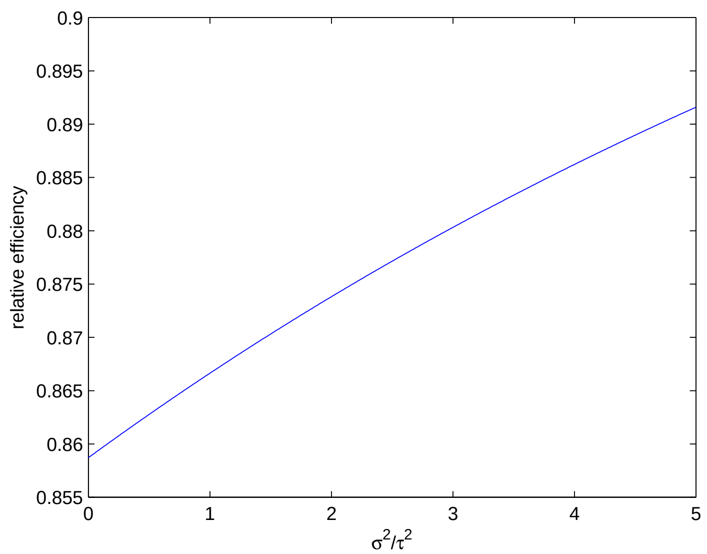


Figure 2: Relative Efficiency of $D_2$ to $D_1$ ($r = 1/3$).

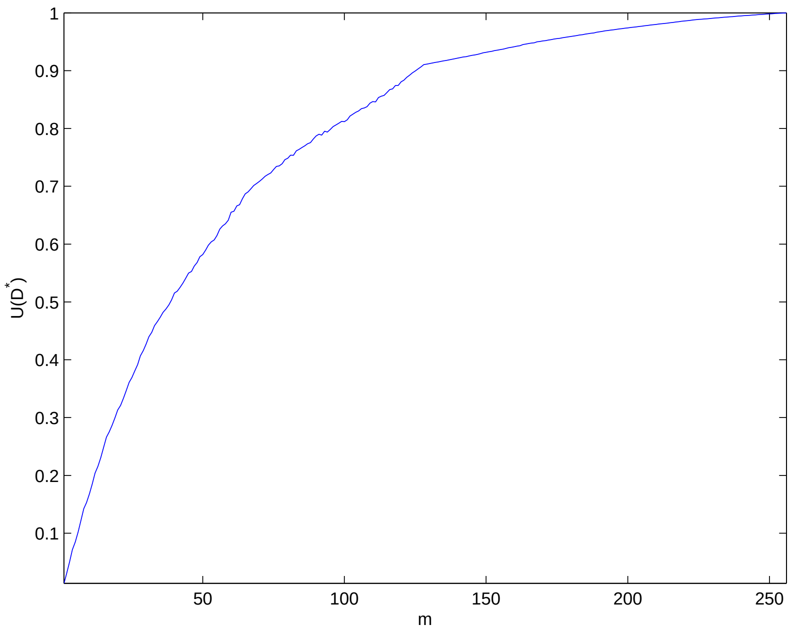


Figure 3: Utility function value of optimal designs with $m = 1, \ldots, 256$ ($r = 1/3$, $\sigma^2/\tau^2 = 0$).

optimal design returned by the algorithm is a fractional factorial design with defining contrast subgroup: $I = ADE = ABC = BCDE$. Furthermore, it can be converted to a cross array because the noise factor does not appear in the defining contrast subgroup. Denote this design by $D_1$.

The Minimum J-aberration single array for this setting given by Wu and Zhu (2003) is $D_2$: $I = ABD = aACE = aBCDE$. It can be shown numerically that $U(D_1) > U(D_2)$ for all $r \in (0, 1)$ and therefore, $D_1$ is uniformly better than $D_2$. We also compared the two designs from a frequentist point of view. Let $N_C$ denote the number of clear control main effects, $N_n$ the number of clear noise main effects, and so on. The estimation capacity of $D_1$ and $D_2$ in terms of clear effects is summarized in Table 3. Although $D_2$ has more number of clear effects, it has three less clear control-by-noise interactions than $D_1$. Therefore, $D_1$ is a much better design for studying robustness.

Table 3: Comparison of Estimation Capacity of $D_1$ and $D_2$ for Example 2.

| Design | Clear Effects | $N_C$ | $N_n$ | $N_{CC}$ | $N_{Cn}$ | $N_{nn}$ |
|---|---|---|---|---|---|---|
| Optimal design ($D_1$) | $a$, $aA$, $aB$, $aC$, $aD$, $aE$ | 0 | 1 | 0 | 5 | 0 |
| Wu-Zhu ($D_2$) | $a$, $C$, $E$, $aB$, $aD$, $BC$, $BE$, $CD$, $DE$ | 2 | 1 | 4 | 2 | 0 |

# 4 MIXED-LEVEL EXPERIMENTS

Experiments with mixed two- and three-level factors are very common in practice. In this section we explain how to design such experiments. The three levels occur only with control factors because we are not interested in choosing an optimum level for a noise factor and therefore, it is enough to study the noise factors using two levels.

Suppose there are $k_n$ two-level noise factors, $k_{C2}$ two-level and $k_{C3}$ three-level control factors. Denote the three levels by $-1$, 0, and 1. As in (8), the transfer function can be written as

$$f(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}) = \mu_0 + \boldsymbol{\beta}'\boldsymbol{u}, \tag{18}$$

where $\boldsymbol{u} = (1, u_1, u_2, \ldots)'$ are functions of $\boldsymbol{x}$ and $\boldsymbol{z}$ representing their main effects and interactions. There are a total of $p = 2^{k_n+k_{C2}}3^{k_{C3}}$ $u$-variables. The main task is to postulate a prior distribution for the parameters that satisfy effect hierarchy and then, to derive the optimal design criterion. The first part is made simple by using the functionally induced priors in Joseph and Delaney (2007).

Note that for two-level experiments, we used the prior

$$\boldsymbol{\beta} \sim N(0, \tau^2\boldsymbol{R}), \quad \text{where } \boldsymbol{R} = diag\{1, r, \ldots, r, r^2, \ldots, r^2, \ldots, r^{k_C+k_n}\}.$$

The prior variance can be equivalently written using Kronecker products as (Joseph and Delaney 2007):

$$var(\boldsymbol{\beta}) = \sigma_0^2 \bigotimes_{j=1}^{k_C+k_n} \boldsymbol{U}_j^{-1}\boldsymbol{\Psi}_j(\boldsymbol{U}_j^{-1})'. \tag{19}$$

where $\sigma_0^2 = \tau^2(1+r)^{k_C+k_n}$, $r = (1-\rho)/(1+\rho)$,

$$\boldsymbol{U}_j = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \text{ and } \boldsymbol{\Psi}_j = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{20}$$

The $\boldsymbol{U}_j$ and $\boldsymbol{\Psi}_j$ are the model matrix and correlation matrix of the $j$th two-level factor. The form in (19) can be easily extended to specify the prior for mixed-level factors. Thus $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \tau^2\boldsymbol{R})$, where

$$\tau^2\boldsymbol{R} = \sigma_0^2 \bigotimes_{j=1}^{k_n} \boldsymbol{U}_j(z_j)^{-1}\boldsymbol{\Psi}_j(z_j)(\boldsymbol{U}_j^{-1}(z_j))' \bigotimes_{j=1}^{k_{C2}+k_{C3}} \boldsymbol{U}_j(x_j)^{-1}\boldsymbol{\Psi}_j(x_j)(\boldsymbol{U}_j(x_j)^{-1})'. \tag{21}$$

The matrices $\boldsymbol{U}_j$ and $\boldsymbol{\Psi}_j$ for two-level factors are given in (20). For a three-level factor the choice of these matrices depends on whether the factor is qualitative or quantitative.

Joseph and Delaney (2007) suggested using Helmert coding for qualitative factors and orthogonal polynomial coding for quantitative factors. For the case of three-level factors, these two coding schemes lead to identical model matrix:

$$\boldsymbol{U}_j = \begin{pmatrix} 1 & -\sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \\ 1 & 0 & -\sqrt{2} \\ 1 & \sqrt{\frac{3}{2}} & \sqrt{\frac{1}{2}} \end{pmatrix}. \tag{22}$$

An isotropic correlation function is recommended for qualitative factors and a Gaussian correlation function for quantitative factors. Thus, the correlation matrices for qualitative and quantitative factors are

$$\boldsymbol{\Psi}_j = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix} \text{ and } \boldsymbol{\Psi}_j = \begin{pmatrix} 1 & \rho & \rho^4 \\ \rho & 1 & \rho \\ \rho^4 & \rho & 1 \end{pmatrix}, \tag{23}$$

respectively.

Let $\boldsymbol{D}$ be the $m\times(k_{C2}+k_{C3}+k_n)$ design matrix and $\boldsymbol{U}_D$ the $m\times p$ model matrix. The following result extends the Proposition 1 to the more general case of mixed-levels.

**Proposition 2.** *If we choose the model matrices as in (20) and (22), then the loss function in* (7) *becomes*

$$L(\boldsymbol{D}) = tr\left(\boldsymbol{A}var(\boldsymbol{\beta}|\boldsymbol{y})\right), \tag{24}$$

*where $\boldsymbol{A}$ is a diagonal matrix whose diagonal entries corresponding to effects containing one and only one noise factor are 1, and 0 otherwise and $var(\boldsymbol{\beta}|\boldsymbol{y})$ is given in (10).*

Define $U(\boldsymbol{D})$ as in (14). As before, we neglect the term $\sigma^2/\tau^2$ in $U(\boldsymbol{D})$ by assuming a high signal-to-noise ratio. We also need to choose a value for $\rho$ for computing $U(\boldsymbol{D})$. For a two-level factor, we know that $r = 1/3$ is a reasonable choice. Thus, from the relation $r = (1-\rho)/(1+\rho)$, we obtain $\rho = 1/2$. This value for $\rho$ will be used for both qualitative and quantitative factors. For the number of runs, we use the same argument as in (15). We obtain $m \geq (1+k_n)(1+k_{C2}+2k_{C3})$.

*Example* 3: Suppose an experiment contains one two-level noise factor $a$, two three-level qualitative control factors $A$ and $B$, and two three-level quantitative control factors $C$ and $D$. The number of runs $m \geq 18$. Therefore, we choose $m = 18$. The optimal design $D_1$ is given in Table 4.

Table 4: The design $D_1$ and $D_3$ for Example 3.

| | $D_1$ | | | | | $D_3$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| run | $A$ | $B$ | $C$ | $D$ | $a$ | $A$ | $B$ | $C$ | $D$ | $a$ |
| 1 | −1 | −1 | −1 | −1 | −1 | −1 | 0 | −1 | 1 | −1 |
| 2 | −1 | −1 | −1 | −1 | 1 | 0 | −1 | 1 | 1 | 1 |
| 3 | 0 | 1 | 0 | −1 | −1 | 1 | 0 | 0 | 1 | 1 |
| 4 | 0 | 1 | 0 | −1 | 1 | 1 | −1 | −1 | 0 | 1 |
| 5 | −1 | 0 | 1 | −1 | −1 | 1 | 1 | 1 | 1 | −1 |
| 6 | −1 | 0 | 1 | −1 | 1 | 0 | 0 | 1 | −1 | −1 |
| 7 | 1 | 0 | −1 | 0 | −1 | 0 | 1 | 0 | 0 | 1 |
| 8 | 1 | 0 | −1 | 0 | 1 | −1 | −1 | 1 | 0 | −1 |
| 9 | 1 | −1 | 0 | 0 | −1 | 1 | 1 | 1 | −1 | 1 |
| 10 | 1 | −1 | 0 | 0 | 1 | 1 | −1 | −1 | −1 | −1 |
| 11 | 0 | −1 | 1 | 0 | −1 | 0 | 0 | −1 | −1 | 1 |
| 12 | 0 | −1 | 1 | 0 | 1 | −1 | −1 | 0 | −1 | 1 |
| 13 | 0 | 1 | −1 | 1 | −1 | −1 | 1 | 0 | −1 | −1 |
| 14 | 0 | 1 | −1 | 1 | 1 | −1 | 1 | −1 | 1 | 1 |
| 15 | −1 | 0 | 0 | 1 | −1 | −1 | 0 | 1 | 0 | 1 |
| 16 | −1 | 0 | 0 | 1 | 1 | 0 | −1 | 0 | 1 | −1 |
| 17 | 1 | 1 | 1 | 1 | −1 | 0 | 1 | −1 | 0 | −1 |
| 18 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | −1 |

Note that the optimal design is a balanced design, but it is not orthogonal. For comparison, consider another two designs. Design $D_2$ is constructed from the famous orthogonal array $OA(18, 2^1 3^7)$ (Taguchi 1987), in which factors $a$ and $A \sim D$ are assigned to the columns $1 \sim 5$. Design $D_3$ is the D-optimal design for the model containing the 18 effects (main effects and two-factor control-by-noise interactions). For $\rho = 1/2$ we obtain, $U(D_2) = 0.2467 < U(D_3) = 0.2569 < U(D_1) = 0.3679$ and therefore, $D_1$ is a much better design than both $D_2$ and $D_3$. In fact, $D_1$ is uniformly better than $D_2$ and $D_3$ for all $\rho \in (0, 1)$.

For the simplicity of the exposition, we have focused only on the mixed two- and three-level factors. The method can be easily generalized to factors with any number of levels using the functionally induced priors in Joseph and Delaney (2007) and following the derivation of Proposition 2.

# 5 FACTORS WITH INTERNAL NOISE

There are factors such as temperature in a heat treatment process and current in a welding process that fluctuate around a nominal value. The nominal values of such factors can be easily controlled and therefore, are considered as control factors, whereas the variations around the nominal values are uncontrollable and are called *internal noise* factors (Taguchi 1987). Internal noise factors arise quite often in experiments, but surprisingly have received little attention in the experimental design literature. We propose two important guiding principles for designing experiments with internal noise factors:

1. Factors with internal noise should be experimented with at least three levels.
2. For factors with internal noise, the interactions among those factors and the interactions with the other control and external noise factors are important.

To explain these principles, consider a factor ($T_1$) with internal noise. We can represent this factor as $T_1 = t_1 + z_1$, where the nominal value $t_1$ is a control factor and $z_1$ is the internal noise factor (Joseph 2008). The first principle is needed, because it will help to estimate the nonlinearity of the response with respect to $T_1$. By exploiting this nonlinearity, we can understand the interactions between $t_1$ and $z_1$ and thus, achieve robustness (note that $T_1^2 = t_1^2 + z_1^2 + 2t_1z_1$ contains the interaction $t_1z_1$). Of course, the nonlinearity can be estimated only if the factor is varied in at least three levels. The need of second principle can be explained as follows. Consider another control factor $x_1$ and an *external noise* factor $z_2$. By entertaining the interaction $T_1x_1$, we can study the interaction between $x_1$ and $z_1$ and by entertaining the interaction $T_1z_2$, we can study the interaction between $t_1$ and $z_2$. Both these interactions are important for achieving robustness. Similarly, the interaction between two factors with internal noise is also important.

Although a factor with internal noise can be conveniently represented as the sum of a control factor and an internal noise factor, the experimental design and the optimal design criterion do not reduce to the cases that we have discussed so far. This is because of the reason that internal noise factors are not systematically varied in a robust parameter design experiment. Thus, a design of experiment with $T_1$ will only specify the values of $t_1$ but not $z_1$. For illustration, consider the following example.

*Example* 4: Let $T_1$ be a factor with internal noise, $x_1$ a two-level control factor, and $z_2$ an external noise factor. The internal noise and external noise are independent and follow $N(0, \sigma_z^2)$. Because $T_1$ has internal noise, we choose three levels for it. The linear and quadratic effects of $T_1$ using orthogonal polynomial coding are

$$T_{1l} = \sqrt{\frac{3}{2}} T_1 \quad \text{and} \quad T_{1q} = \frac{3}{\sqrt{2}}(T_1^2 - \frac{2}{3}).$$

Note that this coding gives us the same model matrix in (22) when $T_1$ takes the values $-1$, 0, and 1. The transfer function is

$$\begin{aligned} f(x_1, T_1, z_2, \boldsymbol{\beta}) = {} & \mu_0 + \beta_0 + \beta_1 T_{1l} + \beta_2 x_1 + \beta_3 z_2 + \beta_4 T_{1l} x_1 + \beta_5 T_{1l} z_2 + \beta_6 x_1 z_2 \\ & + \beta_7 T_{1q} + \beta_8 T_{1q} x_1 + \beta_9 T_{1q} z_2 + \beta_{10} T_{1l} x_1 z_2 + \beta_{11} T_{1q} x_1 z_2. \end{aligned} \tag{25}$$

Substituting $T_1 = t_1 + z_1$ into the model and neglecting the small terms involving $\sigma_z^4$, we can approximate the variance by

$$\begin{aligned} Var(f) \approx {} & \left(\beta_1 \sqrt{\frac{3}{2}} + \beta_4 \sqrt{\frac{3}{2}} x_1 + \beta_7 3\sqrt{2} t_1 + \beta_8 3\sqrt{2} t_1 x_1\right)^2 \sigma_z^2 \\ & + \left(\beta_3 + \beta_5 \sqrt{\frac{3}{2}} t_1 + \beta_6 x_1 + \beta_9 \frac{3}{\sqrt{2}}(t_1^2 - \frac{2}{3}) + \beta_{10} \sqrt{\frac{3}{2}} t_1 x_1 + \beta_{11} \frac{3}{\sqrt{2}}(t_1^2 - \frac{2}{3}) x_1\right)^2 \sigma_z^2, \end{aligned} \tag{26}$$

which is easily seen as equal to:

$$\left(\frac{\partial f(x_1, t_1, \mathbf{0}, \boldsymbol{\beta})}{\partial T_1}\right)^2 \sigma_z^2 + \left(\frac{\partial f(x_1, t_1, \mathbf{0}, \boldsymbol{\beta})}{\partial z_2}\right)^2 \sigma_z^2.$$

The loss function can be derived using the following result. Let $T_j$ be the $j$th factor with internal noise. Define

$$i_j = \begin{cases} 1, & \text{if the } i\text{th effect contains } T_{jl} \text{ or } T_{jq}, \\ 0, & \text{otherwise.} \end{cases} \tag{27}$$

For $i_j = 1$, define

$$\delta_{i,j} = \begin{cases} 1, & \text{if the } i\text{th effect contains } T_{jl}, \\ 0, & \text{if the } i\text{th effect contains } T_{jq}. \end{cases} \tag{28}$$

**Proposition 3.** *Assume there are $k_{C2}$ two-level and $k_{C3}$ three-level control factors, $k_n$ noise factors, and $k_I$ three-level factors with internal noise. The loss function defined in* (7) *is*

$$L(\boldsymbol{D}) = tr(\boldsymbol{A}var(\boldsymbol{\beta}|\boldsymbol{y})),$$

*where $\boldsymbol{A}$ is a diagonal matrix, whose diagonal entries are given by*

$$A_{ii} = \begin{cases} 1, & \textit{if the effect contains one and only one n, i.e., } n, Cn, T_l n, \ldots \\ \sum_{j=1}^{k_I} i_j 12/8^{\delta_{i,j}}, & \textit{if the effect contains } T \textit{ but no } n\textit{, i.e., } T_l, T_q, CT_l, CT_q, \ldots \\ 0, & \textit{otherwise.} \end{cases} \tag{29}$$

Consider again the Example 4. Using Proposition 3, we obtain the loss function as

$$L(\boldsymbol{D}) = \{\frac{3}{2}V_1 + \frac{3}{2}V_4 + 12V_7 + 12V_8 + V_3 + V_5 + V_6 + V_9 + V_{10} + V_{11}\}.$$

For the frequentist estimation of $x_1$, $z_2$, $T_{1l}$, $T_{1q}$, $x_1z_2$, $T_{1l}z_2$, $T_{1l}x_1$, and the grand mean, $m \geq 8$. Therefore, we choose $m = 8$. Using the same definition for $U(D)$ as before, we can directly apply our exchange algorithm here as long as matrix $\boldsymbol{A}$ is specified correctly. The optimal design $D_1$ is given in Table 5. Interestingly, the optimal design happened to be an orthogonal main effects plan $OME(8, 3^1 2^2)$, which can be produced by collapsing the two middle levels of the four-level factor in $OA(8, 4^1 2^2)$ into level 0 (Wu and Hamada 2000, Section 7.8). For comparison, suppose we construct another $OME(8, 3^1 2^2)$ by collapsing the highest two levels of the four-level factor into level 1. This design $D_2$ is also shown in Table 5. It can be shown that $U(D_1) > U(D_2)$ for all $\rho \in (0, 1)$ and therefore, $D_1$ is a better design than $D_2$. This could be because $D_1$ can estimate the quadratic effect of $T_1$ better and thus, can obtain a better estimate of the control-by-noise interaction $t_1z_1$ than $D_2$.

# 6 CONCLUSIONS

In this article we proposed a Bayesian optimal design criterion for constructing single arrays for robust parameter design experiments. Different from others, the Bayesian criterion is capable

Table 5: $D_1$ and $D_2$ for Example 4.

| | $D_1$ | | | $D_2$ | | |
|---|---|---|---|---|---|---|
| run | $x_1$ | $z_2$ | $t_1$ | $x_1$ | $z_2$ | $t_1$ |
| 1 | −1 | −1 | −1 | −1 | −1 | −1 |
| 2 | 1 | 1 | −1 | 1 | 1 | −1 |
| 3 | −1 | 1 | 0 | −1 | 1 | 0 |
| 4 | 1 | −1 | 0 | 1 | −1 | 0 |
| 5 | −1 | −1 | 0 | −1 | −1 | 1 |
| 6 | 1 | 1 | 0 | 1 | 1 | 1 |
| 7 | −1 | 1 | 1 | −1 | 1 | 1 |
| 8 | 1 | −1 | 1 | 1 | −1 | 1 |

of incorporating the importance of control-by-noise interactions without altering the effect hierarchy principle. Several examples show the superiority of the proposed single arrays; in both Bayesian and frequentist viewpoints.

We have made several assumptions to simplify the approach. Specifically, we have chosen to fix $r = 1/3$ and $\sigma^2/\tau^2 = 0$ for finding the optimal single arrays. A better approach could be to postulate a second stage prior on these hyperparameters and give a fully Bayesian treatment, but this will be at the cost of increased computations. Many examples that we tried so far show that the foregoing choices of $r$ and $\sigma^2/\tau^2$ are quite reasonable. Moreover, in many cases the single arrays we obtained are uniformly optimal for all choices of these hyperparameters.

It is also possible to incorporate effect heredity principle (see, e.g., Wu and Hamada 2000) in our Bayesian optimal design criterion. This can be done using a different $\rho_i$ for each factor. See Joseph and Delaney (2007) for the details of this prior specification. However, rarely those values will be known before the experiment and therefore, a fully Bayesian treatment is warranted. We leave this as a topic for future research.

## ACKNOWLEDGMENTS

This research was supported by U.S. National Science Foundation grant CMMI-0448774. We thank the Associate Editor for the valuable comments and suggestions.

## APPENDIX: PROOFS

### Proof of (16)

We have:

$$
\begin{aligned}
\tilde{\boldsymbol{U}}_D \boldsymbol{R} \tilde{\boldsymbol{U}} &= \begin{pmatrix} \boldsymbol{U}_D \\ \boldsymbol{F}' \end{pmatrix} \boldsymbol{R} \left(\boldsymbol{U}_D', \boldsymbol{F}\right) \\
&= \begin{pmatrix} \boldsymbol{U}_D \boldsymbol{R} \boldsymbol{U}_D', & \boldsymbol{U}_D \boldsymbol{R} \boldsymbol{F} \\ \boldsymbol{F}' \boldsymbol{R} \boldsymbol{U}_D', & \boldsymbol{F}' \boldsymbol{R} \boldsymbol{F} \end{pmatrix}
\end{aligned}
$$

and thus,

$$
(\tilde{\boldsymbol{U}}_D \boldsymbol{R} \tilde{\boldsymbol{U}}_D')^{-1} = \frac{1}{d} \begin{pmatrix} d(\boldsymbol{U}_D \boldsymbol{R} \boldsymbol{U}_D')^{-1} + \boldsymbol{H}' \boldsymbol{F} \boldsymbol{F}' \boldsymbol{H}, & -\boldsymbol{H}' \boldsymbol{F} \\ -\boldsymbol{F}' \boldsymbol{H}, & 1 \end{pmatrix},
$$

where $d = \boldsymbol{F}'(\boldsymbol{R} - \boldsymbol{M})\boldsymbol{F}$, $\boldsymbol{H} = \boldsymbol{R}\boldsymbol{U}_D'(\boldsymbol{U}_D \boldsymbol{R} \boldsymbol{U}_D')^{-1}$, and $\boldsymbol{M} = \boldsymbol{R}\boldsymbol{U}_D'(\boldsymbol{U}_D \boldsymbol{R} \boldsymbol{U}_D')^{-1}\boldsymbol{U}_D \boldsymbol{R} = \boldsymbol{H}\boldsymbol{U}_D \boldsymbol{R}$. Hence, we obtain (16):

$$
\begin{aligned}
\boldsymbol{M}_x &= (\boldsymbol{R}\boldsymbol{U}_D', \boldsymbol{R}\boldsymbol{F}) \frac{1}{d} \begin{pmatrix} d(\boldsymbol{U}_D \boldsymbol{R} \boldsymbol{U}_D')^{-1} + \boldsymbol{H}' \boldsymbol{F} \boldsymbol{F}' \boldsymbol{H}, & -\boldsymbol{H}' \boldsymbol{F} \\ -\boldsymbol{F}' \boldsymbol{H}, & 1 \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_D \boldsymbol{R} \\ \boldsymbol{F}' \boldsymbol{R} \end{pmatrix} \\
&= \boldsymbol{M} + \frac{1}{d} \left((\boldsymbol{R} - \boldsymbol{M})\boldsymbol{F}\right) \left((\boldsymbol{R} - \boldsymbol{M})\boldsymbol{F}\right)'.
\end{aligned}
$$

### Proof of Propositions 1 and 2

Because Proposition 1 is a special case of Proposition 2 when $k_{C3} = 0$ and $k_{C2} > 0$, we only need to prove Proposition 2. We can write the transfer function (18) as

$$
\begin{aligned}
f(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\beta}) = \mu_0 &+ \sum_{i=0}^{N-1} \beta_i^0 v_i + \left\{ \sum_{i=0}^{N-1} \beta_i^1 v_i \right\} z_1 + \cdots + \left\{ \sum_{i=0}^{N-1} \beta_i^{k_n} v_i \right\} z_{k_n} \\
&+ \left\{ \sum_{i=0}^{N-1} \beta_i^{12} v_i \right\} z_1 z_2 + \cdots + \left\{ \sum_{i=0}^{N-1} \beta_i^{12\ldots k_n} v_i \right\} z_1 \ldots z_{k_n},
\end{aligned} \tag{30}
$$

where $z_i$ are noise factors, $v_i$'s are generic notations for the linear, quadratic, and interaction effects among control factors, and $N = 2^{k_{C2}} 3^{k_{C3}} = |\mathscr{X}|$. Let $x_{kl}$ and $x_{kq}$ denote the linear and quadratic effects of $x_k$ defined based on the coding in (22). Note that because the coding is the same for a qualitative factor, for simplicity, we use the same terms "linear" and "quadratic" effects to refer to its two components. Then, we can write $v_i$ in the form

$$
v_i = \prod_{k=1}^{k_{C2}} x_k^{\gamma_k} \prod_{k=1}^{k_{C3}} x_{kl}^{\min\{\alpha_{i,k}, \gamma_k\}} x_{kq}^{\min\{1-\alpha_{i,k}, \gamma_k\}}. \tag{31}
$$

Because of the coding scheme, we obtain

$$\text{for a two-level factor:} \quad \sum_{x_k=-1,1} 1 = \sum_{x_k=-1,1} x_k^2 = 2 \tag{32}$$

$$\text{for a three-level factor:} \quad \sum_{x_k=-1,0,1} 1 = \sum_{x_k=-1,0,1} x_{kl}^2 = \sum_{x_k=-1,0,1} x_{kq}^2 = 3.$$

Define $\gamma_k = 1$ if $v_i$ contains $x_k$ and 0 otherwise. For a three-level factor $x_k$, if $\gamma_k = 1$, define $\alpha_{i,k} = 1$ if $v_i$ contains $x_{kl}$ and $\alpha_{i,k} = 0$ if $v_i$ contains $x_{kq}$. Note that $v_i$ cannot contain both $x_{kl}$ and $x_{kq}$ for the same $x_k$.

From the transfer function, we obtain

$$d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}}) = \left( \sum_{i=0}^{N-1} (\beta_i^1 - \hat{\beta}_i^1) v_i, \ldots, \sum_{i=0}^{N-1} (\beta_i^{k_n} - \hat{\beta}_i^{k_n}) v_i \right)'. \tag{33}$$

Apparently, $\sum_{\boldsymbol{x} \in \mathscr{X}} v_i v_j = 0$ if $i \neq j$ and $\sum_{\boldsymbol{x} \in \mathscr{X}} v_i^2 = 2^{k_{C2}} 3^{k_{C3}}$. Therefore,

$$\begin{aligned} L(\boldsymbol{D}) &= \frac{1}{|\mathscr{X}|} \sum_{\boldsymbol{x} \in \mathscr{X}} E\{d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}})' d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}})\} \\ &= \frac{1}{|\mathscr{X}|} E\{\sum_{\boldsymbol{x} \in \mathscr{X}} d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}})' d(\boldsymbol{x}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}})\} \\ &= E\{\sum_{j=1}^{k_n} \sum_{i=0}^{N-1} (\beta_i^j - \hat{\beta}_i^j)^2\} \end{aligned}$$

Because $\hat{\boldsymbol{\beta}} = E(\boldsymbol{\beta}|\boldsymbol{y})$, $E\{(\beta_i^j - \hat{\beta}_i^j)^2\} = EE\{(\beta_i^j - \hat{\beta}_i^j)^2|\boldsymbol{y}\} = E\{var(\beta_i^j|\boldsymbol{y})\}$. Now, by the property of normal distribution $var(\boldsymbol{\beta}|\boldsymbol{y})$ is independent of $\boldsymbol{y}$. Thus, we obtain

$$L(\boldsymbol{D}) = \sum_{j=1}^{k_n} \sum_{i=0}^{N-1} var(\beta_i^j|\boldsymbol{y}) = tr(\boldsymbol{A} var(\boldsymbol{\beta}|\boldsymbol{y})),$$

where $\boldsymbol{A}$ follows the definition in Proposition 1 and 2.

## Proof of Proposition 3

Let $T_k$ be a factor with internal noise and $t_k$ its nominal value. Using orthogonal polynomial coding, $T_{kl}$ and $T_{kq}$ are the linear and quadratic effects of $T_k$. Accordingly, $t_{kl} = \sqrt{3/2} t_k$ and $t_{kq} = 3/\sqrt{2}(t_k^2 - 2/3)$. In this proof, $\mathscr{X}$ is the design space consisting of all the control factors $x_k$, $k = 1, \ldots, k_{C2} + k_{C3}$, and $t_k$, $k = 1, \ldots, k_I$. Write the transfer function in the same form as in

(30). Let $v_i$ be the effects involving only control factors and factors with internal noise. We can write $v_i$ as:

$$v_i = \prod_{k=1}^{k_{C2}} x_k^{\gamma_k} \prod_{k=1}^{k_{C3}} x_{kl}^{\min\{\alpha_{i,k},\gamma_k\}} x_{kq}^{\min\{1-\alpha_{i,k},\gamma_k\}} \prod_{k=1}^{k_I} T_{kl}^{\min\{\delta_{i,k},i_k\}} T_{kq}^{\min\{1-\delta_{i,k},i_k\}}, \tag{34}$$

where $\gamma_k$ and $\alpha_{i,k}$ are defined in the proof of Proposition 2 and $i_k$ and $\delta_{i,k}$ are defined in (27) and (28). Let $N = 2^{k_{C2}} 3^{k_{C3}+k_I} = |\mathscr{X}|$.

Now, let us study an element of $d(\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{\beta}, \widehat{\boldsymbol{\beta}})$ (use $d$ for short ). When

$$d_j = \frac{\partial f(\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{0}, \boldsymbol{\beta})}{\partial z_j} - \frac{\partial f(\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{0}, \widehat{\boldsymbol{\beta}})}{\partial z_j} = \sum_{i=0}^{N-1} (\beta_i^j - \hat{\beta}_i^j) v_i,$$

where $v_i$ is the same as in (34) except that $T_{kl}$ and $T_{kq}$ are replaced by the nominal values $t_{kl}$ and $t_{kq}$ for $k = 1, \ldots, k_I$. Then we have $\sum_{\mathscr{X}} v_i v_j = 0$ and $\sum_{\mathscr{X}} v_i^2 = N$, and hence

$$\frac{1}{|\mathscr{X}|} \sum_{\mathscr{X}} \int d_j^2 p(\boldsymbol{\beta}, \boldsymbol{y}) d\boldsymbol{\beta} d\boldsymbol{y} = \sum_{i=0}^{N-1} var(\beta_i^j | \boldsymbol{y}).$$

When

$$d_j = \frac{\partial f(\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{0}, \boldsymbol{\beta})}{\partial T_j} - \frac{\partial f(\boldsymbol{x}, \boldsymbol{t}, \boldsymbol{0}, \widehat{\boldsymbol{\beta}})}{\partial T_j} = \sum_{i=0}^{N-1} (\beta_i^0 - \hat{\beta}_i^0) \frac{\partial v_i}{\partial T_j},$$

define

$$b_j = \prod_{k=1}^{k_{C2}} x_k^{\gamma_k} \prod_{k=1}^{k_{C3}} x_{kl}^{\min\{\alpha_{i,k},\gamma_k\}} x_{kq}^{\min\{1-\alpha_{i,k},\gamma_k\}} \prod_{k \neq j}^{k_I} t_{kl}^{\min\{\delta_{i,k},i_k\}} t_{kq}^{\min\{1-\delta_{i,k},i_k\}}.$$

Note that $\sum_{\mathscr{X}} b_j^2 = \sum_{\mathscr{X}} b_j^2 t_{jl}^2 = N$. There are only three cases for $v_i$:

1. If $i_j$ for $T_j$ is 0, then $\frac{\partial v_i}{\partial T_j} = 0$.
2. If $i_j = 1$ and $\delta_{i,j} = 1$, then $\frac{\partial v_i}{\partial T_j} = \sqrt{\frac{3}{2}} b_j$. Therefore, $\sum_{\mathscr{X}} \left( \frac{\partial v_i}{\partial T_j} \right)^2 = \frac{3}{2} N$.
3. If $i_j = 1$ and $\delta_{i,j} = 0$, then $\frac{\partial v_i}{\partial T_j} = 3\sqrt{2} t_j b_j = 2\sqrt{3} t_{jl} b_j$. Therefore, $\sum_{\mathscr{X}} \left( \frac{\partial v_i}{\partial T_j} \right)^2 = 12N$.

It is easy to see that in all the three cases $\sum_{\mathscr{X}} \frac{\partial v_i}{\partial T_j} \frac{\partial v_{i'}}{\partial T_j} = 0$ if $i \neq i'$. Hence

$$\begin{aligned} \frac{1}{|\mathscr{X}|} \sum_{\mathscr{X}} \int d_j^2 p(\boldsymbol{\beta}, \boldsymbol{y}) d\boldsymbol{\beta} d\boldsymbol{y} &= \sum_{i=0}^{N-1} \left\{ i_j \left( \frac{3}{2} \right)^{\delta_{i,j}} \times 12^{1-\delta_{i,j}} \right\} var(\beta_i^0 | \boldsymbol{y}) \\ &= \sum_{i=0}^{N-1} \left\{ 12 i_j 8^{-\delta_{i,j}} \right\} var(\beta_i^0 | \boldsymbol{y}). \end{aligned}$$

At last, the loss function is computed as follows:

$$
\begin{aligned}
L(\boldsymbol{D}) &= \frac{1}{|\mathscr{X}|}\sum_{j=1}^{k_l}\sum_{\mathscr{X}}\int d_j^2 p(\boldsymbol{\beta},\boldsymbol{y})d\boldsymbol{\beta}d\boldsymbol{y} \\
&= \sum_{i=0}^{N-1}\left\{\sum_{j=1}^{k_l} 12 i_j 8^{-\delta_{i,j}}\right\} var(\beta_i^0|\boldsymbol{y}) + \sum_{j=1}^{k_n}\sum_{i=0}^{N-1} var(\beta_i^j|\boldsymbol{y}) \\
&= tr(\boldsymbol{A} var(\boldsymbol{\beta}|\boldsymbol{y})).
\end{aligned}
$$

Thus, the diagonal entries of $\boldsymbol{A}$ corresponding to $\beta_i^0$ is $\sum_{j=1}^{k_l} 12 i_j 8^{-\delta_{i,j}}$, corresponding to $\beta_i^j$ is 1, and all others are 0.

# REFERENCES


Bingham, D., and Sitter, R. R. (2003), "Fractional Factorial Split-Plot Designs for Robust Parameter Experiments," *Technometrics*, 45, 80-89.

Borror, C. M., Montgomery, D. C., and Myers, R. H. (2002), "Evaluation of Statistical Designs for Experiments Involving Noise Variables," *Journal of Quality Technology*, 34, 54-70.

Del Castillo, E., Alvarez, M. J., Ilzarbe, L., and Viles, E. (2007), "A New Design Criterion for Robust Parameter Experiments," *Journal of Quality Technology*, 39, 279-295.

Ginsburg, H., and Ben-Gal, I. (2006), "Designing experiments for robust-optimization problems: the Vs-optimality criterion," *IIE Transactions on Quality and Reliability Engineering*, 38, 445-461.

Joseph, V. R. (2006), "A Bayesian Approach to the Design and Analysis of Fractionated Experiments," *Technometrics*, 48, 219-229.

Joseph, V. R. (2007), "Taguchi's Approach to Robust Parameter Design: A New Perspective," *IIE Transactions on Quality and Reliability Engineering*, 39,805-810.

Joseph, V. R. (2008), "Tolerance Design". *Encyclopedia of Statistics in Quality and Reliability* (Eds. Ruggeri, F., Kennet, R. S., and Faltin, F.), New York: Wiley, 2014-2019.

Joseph, V. R., and Delaney, J. D. (2007), "Functionally Induced Priors for the Analysis of Experiments," *Technometrics*, 49, 1-11.

Li, W. W., and Wu, C. F. J. (1997), "Columnwise-Pairwise Algorithms With Applications to the Construction of Supersaturated Designs," *Technometrics*, 39, 171-179.

Li, X., Sudarsanam, N., and Frey, D. D. (2006), "Regularities in Data from Factorial Experiments," *Complexity*, 11, 32-45.

Meyer, R. K., and Nachtsheim, C. J. (1995), "The Coordinate-Exchange Algorithm for Constructing Exact Optimal Experimental Designs," *Technometrics*, 37, 60-69.

Miller, A., Sitter, R., Wu, C. F. J., and Long, D. (1993), "Are Large Taguchi-style Experiments Necessary? A Reanalysis of Gear and Pinion Data," *Quality Engineering*, 6, 21-37.

Nguyen, N., and Miller, A. J. (992), "A Review of Some Exchange Algorithm for Constructing Discrete D-optimal Designs," *Computational Statistics & Data Analysis*, 14, 489-498.

Shoemaker, A. C., Tsui, K. L. and Wu, C. F. J. (1991), "Economical experimentation methods for robust design." *Technometrics*, 33, 415-427.

Taguchi, G. (1987), *System of Experimental Design*, Vol 1 & 2, White Plains, New York: Unipub/Kraus International.

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.

Wu, C. F. J., and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: Wiley.

Wu, C. F. J., and Zhu, Y. (2003), "Optimal Selection of Single Array for Parameter Design Experiments," *Statistica Sinica*, 13, 1179-1199.

Zhu, Y., Zeng, P., and Jennings, K. (2007), "Optimal Compound Orthogonal Arrays and Single Arrays for Robust Parameter Design Experiments," *Technometrics*, 49, 440-453.