System Regularities in Design of Experiments and Their Application

by

Xiang Li

B.S. and M.S., Engineering Mechanics
Tsinghua University, 2000, 2002

Submitted to the Department of Mechanical Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Mechanical Engineering

at the

Massachusetts Institute of Technology

June, 2006

Signature of Author …………………………………………………………….....
Department of Mechanical Engineering
May 9, 2006

Certified by …………………………………………………………………………
Daniel D. Frey
Assistant Professor of Mechanical Engineering and Engineering Systems
Thesis Supervisor

Accepted by ..……………………………………………………………………..
Lallit Anand
Chairman, Department Committee on Graduate Students

System Regularities in Design of Experiments and Their Application

by

Xiang Li

Submitted to the Department of Mechanical Engineering on May 9, 2006,
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy in Mechanical Engineering

ABSTRACT

This dissertation documents a meta-analysis of 113 data sets from published factorial experiments. The study quantifies regularities observed among main effects and multi-factor interactions. Such regularities are critical to efficient planning and analysis of experiments, and to robust design of engineering systems. Three previously observed properties are analyzed – effect sparsity, hierarchy, and heredity. A new regularity on effect synergism is introduced and shown to be statistically significant. It is shown that a preponderance of active two-factor interaction effects are synergistic, meaning that when main effects are used to increase the system response, the interactions provide an additional increase and that when main effects are used to decrease the response, the interactions generally counteract the main effects.

Based on the investigation of system regularities, a new strategy is proposed for evaluating and comparing the effectiveness of robust parameter design methods. A hierarchical probability model is used to capture assumptions about robust design scenarios. A process is presented employing this model to evaluate robust design methods. This process is then used to explore three topics of debate in robust design: 1) the relative effectiveness of crossed versus combined arrays; 2) the comparative advantages of signal-to-noise ratios versus response modeling for analysis of crossed arrays; and 3) the use of adaptive versus "one shot" methods for robust design. For the particular scenarios studied, it is shown that crossed arrays are preferred to combined arrays regardless of the criterion used in selection of the combined array. It is shown that when analyzing the data from crossed arrays, signal-to-noise ratios generally provide superior performance; although that response modeling should be used when three-factor interactions are absent. Most significantly, it is shown that using an adaptive inner array design crossed with an orthogonal outer array resulted in far more improvement on average than other alternatives.

Thesis Supervisor: Daniel D. Frey
Title: Assistant Professor of Mechanical Engineering and Engineering Systems

Acknowledgments

I would like to thank a number of people who contributed in many ways in making this thesis possible. First, I am grateful to my advisor, Professor Daniel D. Frey, for giving me the opportunity to work on such a challenging topic. With deep research insights and rich engineering knowledge, he has patiently guided me to explore the complicated topic of effectively applying experimental designs to engineering systems and to conquer a variety of difficulties along the way. In addition, he always gave me the freedom to pursue my interests and always offered me great opportunities to widen my perspective. I also wish to thank my doctoral committee members, Professor Warren P. Seering and Professor Roy E. Welsch, for their invaluable advice and inspiring discussions on this research.

Thanks to the staff members of MIT who have been in great assistance to this work. I would especially like to thank Ms. Leslie Regan, Ms. Joan Kravit, Ms. Maggie Sullivan, Mr. Jason Pring, and Ms. Danielle Guichard-Ashbrook. I also owe a debt of gratitude to Mr. Steven Schondorf for being my mentor and giving me continuous support.

I would like to sincerely acknowledge the encouragement from my colleagues in the Robust Design Group: Rajesh Jugulum, Yiben Lin, Hungjen Wang, Chad Forster, Nandan Sudarsanam, Danielle Zurovcik, Jagmeet Singh, and Troy Savoie, who have been wonderful friends and research partners.

Finally, I would like to express my deepest thanks to my parents and my sister for their love and encouragement. This work could not have been done without their support. In addition, my close friends have always kept me sane and motivated: Elaine, Chen, Shuang, Dave, Bolero, Huajie, and Hao – thanks for everything.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Experimentation is an important activity in design of engineering systems[1]. Most existing engineering systems were shaped by a process of experimentation including preliminary investigation of phenomena, sub-system prototyping, and system verification tests. Major, complex systems typically require thousands of experiments [1].

The process of experimental design (or Design of Experiments, DOE) uses statistical techniques to organize experiments so that the right type of data, and enough of it, is available to answer the questions of interest as clearly and efficiently as possible.

With the fast development of modern computers and CAD/CAE software, more and more experimental designs and relative studies are carried out on computers using Monte Carlo simulations, Bayesian analysis, and other computational techniques. Model accuracy becomes a crucial issue in the Design of Experiments. If a model does not accurately reflect system properties, the validity of predictions and decisions made based on this model will certainly suffer. Therefore, the ability to build sound models for systems under investigation and to assign

[1] Engineering systems in this dissertation refer to devices, products and processes used in engineering design, manufacturing, and other activities.

proper values to the parameters in those models is in urgent demand. Half of this dissertation focuses on this aspect.

In the DOE literature, three are most documented system regularities are Effect Sparsity, Effect Hierarchy, and Effect Heredity. These regularities play a vital role in both experimental design theory and practice. Research performed with Professor Frey found that they could essentially affect the effectiveness of DOE and Robust Parameter Design (RPD) methods [2]. However, these regularities were not quantified. They were incorporated into DOE models using heuristic values based on the experimenters' experience. With the current booming of modeling and analyzing complex engineering systems on computers, the verification and quantification of these regularities are urgently needed. I collected more than a hundred engineering experimental data sets from a large range of engineering subjects and built an engineering experiment database. Based on this database, all three regularities were verified and quantified. The results offer quantitative proof of the existence of those three regularities and suggested caution on their use.

One of the important issues in industrial experiments is how to make experiments more efficient and reliable. New experimental design techniques have been developed to meet these goals and they generally rely on system regularities. To search for new regularities in engineering systems offers the potential to improve current experimental designs or develop new designs. Also, it helps to reach a deeper understanding of the results of applying DOE in industry. Therefore, we dig into the database and in addition to the current regularities, we identified a new regularity in the effect interactions which we named Asymmetric Synergistic Interaction Structure (ASIS). It has potential to improve current models in DOE and make an impact on the Bayesian model selection method and other DOE strategies.

As an important approach to enhance system robustness, Robust Parameter Design (RPD) has become a center topic in lots of system design curricula. Another major focus of this dissertation is robust parameter designs. Robust Parameter Design is a set of design methods in which products and processes are made less sensitive to manufacturing variations, customer use conditions, and degradation over time. In practice, it is very difficult to compare different robust design methods. Because experimentation usually costs a large amounts of time and money, practitioners do not carry out experiments comparing different robust design methods. On the other hand, it is very important to understand the performance of different RPD methods so that we can accurately estimate budget-over-improvement ratios, manage the RPD process, and understand its results. Using our engineering experiment database and quantification of existing system regularities, we build a hierarchical probability model which represents general engineering systems. With this hierarchical probability model, a new strategy for validating and comparing RPD methods has been generated and implemented. This enabled us to make comparison of different RPD methods directly based on computer simulations. These analysis and applications complete the second half of this dissertation.

1.2 Motivation

There are multiple open questions, academic debates, and recent developments in both the theory and practice of Design of Experiments and Robust Parameter Design which motivate my work. To summarize, my work was strongly motivated by two debates within academia, and by the difference between theoretical analysis and practical observation.

1.2.1 Debate on Adaptive Experimentation

The first debate is on adaptive experimentation. The focus of the debate is whether One-Factor-at-A-Time (OFAT) is a good experimental strategy compared to factorial designs. There are a series of discouragements to using OFAT in experimental designs and quality improvement in the literature. Some of the representative arguments can be found in Box, et al (1978) [3], Logothetis, et al (1994) [4], Czitrom (1999) [5], and Wu, et al (2000) [6]. There are five major reasons cited to discourage using OFAT:

1) OFAT requires more runs for the same precision of effect estimation;
2) OFAT cannot estimate some interactions;
3) Conclusions from OFAT analysis may not be general;
4) OFAT can miss optimal settings of factors;
5) OFAT can be susceptible to bias due to time trends;

On the other hand, there are also a series of supports that demonstrated OFAT had advantages over factorial designs under certain conditions. These arguments can be found in Friedman, et al (1947) [7], Daniel (1973) [8], Koita (1994) [9], McDaniel et al (2000) [10], and Frey et al (2003) [11]. The reasons cited to support using of OFAT can be summarized as below:

1) OFAT might be used in preference to balanced factorial plans within a system likely to contain interactions;
2) OFAT concentrates observations in regions that are likely to contain the optimum;
3) OFAT reacts more quickly to data;
4) OFAT performs better when experimental error is small;

1.2.2 Debate on Robust Design Methods

The second debate is on Robust Design Methods. The focus of the debate is to compare two RPD methods – Cross Arrays[2] and Combined Arrays[3]. Cross Array method was proposed by Taguchi, et al (1985) [12] through their work primarily within industry in Japan. Shortly after it was introduced to North America and Europe, Taguchi's method generally became a subject of scholarly scrutiny. Some examples include Kacker (1985) [13], Hunter (1985) [14], Box (1985, 1988) [15, 16], Hamada, et al (1990) [17], and Nair (1992) [18]. Welch et al (1990) [19] proposed that cross arrays be replaced by combined arrays and this approach was further expanded by Shoemaker, et al (1991) [20], and Borror et al (2000) [21], etc. Wu and Hamada (2000) [6] claimed that combined arrays achieve improved run size economy (or provide advantages in resolving selected effects) over cross arrays. They claim that "Some of the combined arrays are uniformly better than cross arrays of the same run size in terms of number of clear main effects and two-factor interactions" [6].

On the other hand, there is practical evidence that supports cross arrays over combined arrays. Kunert, et al (2005) [22] studied a sheet metal spinning experiment in which he compared the effectiveness of combined arrays and cross arrays. Both methods were applied to a single engineering system and a key result of the study was that the cross arrays outperformed combined arrays by reducing the variance of the response. The authors suggest that cross arrays should be used to provide robustness against imprecise model assumptions.

---

[2] Cross Arrays are sometimes referred as Crossed Arrays or Product Arrays.
[3] Combined Arrays are sometimes referred as Single Arrays.

1.3 Research Objectives

In my understanding, the reason the previous two debates existed for several decades is that there is a gap between theoretical analysis and practical observations. In theoretical analysis, researchers employ models and apply advanced statistical knowledge and techniques to analyze experimental designs. However, models cannot reflect every aspect within a system and we must ensure that we captured the most important features of a system in the models. In practical observations, practitioners carry out hundreds of experiments but the results vary case by case. It is difficult for practitioners buried in daily company projects to summarize all the cases and condense feedback to the academic community. This gap impeded the development of DOE and Robust Design methodology.

My work fills this gap and enables an information flow between theoretical analysis and practical implementations. This can be illustrated in Figure 1.

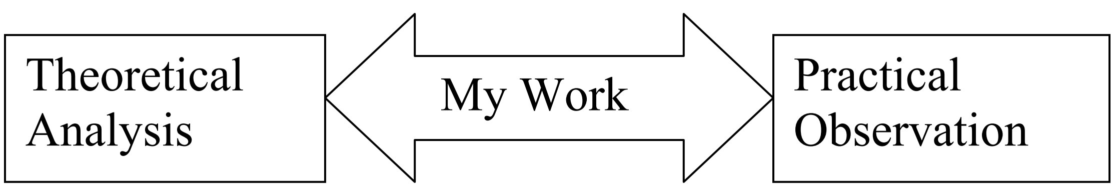


Figure 1 Objectives of my research work

In details, there are several aims to achieve through this work:

1) To collect engineering experimental data sets and build up a large engineering experiment database.

2) To study regularities in engineering systems and their impacts on DOE and RPD methods. To verify and quantify system regularities using the engineering experiment database.
3) To develop general engineering system models that incorporate system regularities and to carry out computer simulations based on these model.
4) To validate and compare different Robust Design methods using the models developed, and to gain a deeper understanding of RPD and recommend efficient and effective methods to robust design practitioners.
5) To search for new regularities in engineering systems, identify and quantify these regularities and articulate their potential impacts on the DOE and RPD methodology.

A broader impact of this research is to influence both the theory and the industry practice of DOE and RPD through a study combining industry experimental cases, statistical models, computer simulations, and quantitative analysis.

1.4 Research Roadmap

My research centers on DOE regularities and their application to robust design. A roadmap of my research is shown in Figure 2. A typical scientific or engineering learning process contains three steps: Observing, Modeling, and Applying. My research process can also be fitted into this framework.

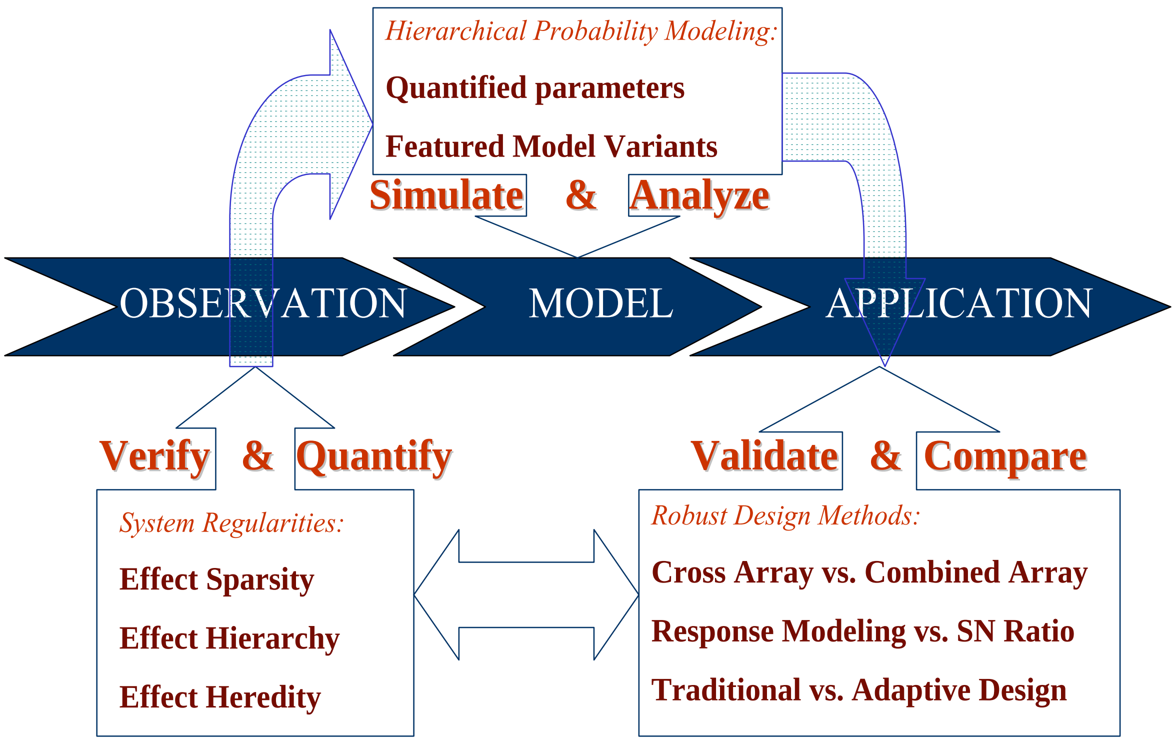


Figure 2 Research roadmap

Three engineering system regularities have been verified and quantified which are discussed in the Design of Experiments literature. These regularities are shown to essentially affect robust designs. The verification and quantification are based on an engineering experiment database which contains more than one hundred engineering experimental data sets collected from published experiments. These results enable the development of a Hierarchical Probability Model which employs all quantified parameters from the former quantification step and can easily generate featured model variants for computer simulation. A new model-based evaluation and comparison strategy is proposed for robust design methods. This strategy offers a direct way to compare robust design methods quantitatively and a more accurate way to estimate the benefit/cost ratio of these methods over empirical estimations. A further analysis comparing robust design methods using this strategy suggests that the comparison results depend on system regularities present in the target systems.

1.5 Organization of the Dissertation

This dissertation is organized according to research topics and follows a natural logic of the research process. There are six chapters in total.

The first chapter gives an overview of my work and introduces the motivation of this research. In this chapter, research objectives are also defined and a research roadmap with detailed explanation is presented.

The second chapter goes into details of system regularities in engineering experiments. It defines and explains the regularities from a Design of Experiments view and why the regularities are important to practitioners. Examples are given in the DOE background introduction and in the system regularity illustrations.

The third chapter describes the approach of verifying and quantifying three regularities – Effect Sparsity, Hierarchy, and Heredity. A meta-analysis of 113 responses from published experiments was carried out and the three widely discussed regularities were tested on these cases. The regularities are verified, but my results suggest using them with caution.

The fourth chapter goes to the application section of this research – applying system regularities to simulations. The goal is to validate and compare different robust design methods and offering practitioners suggestions on choosing efficient robust designs. Two model-based comparisons were carried out, one to compare single array methods with cross array methods and the other to compare adaptive methods with traditional methods. The results generally show benefit in cross array methods over adaptive methods. Benefit is also shown in a new approach that incorporates adaptive methods in cross arrays and the new approach renders a great improvement over traditional methods.

The fifth chapter is devoted to the Asymmetric Synergistic Interaction Structure (ASIS) which is a new regularity observed and quantified in the engineering experiment database. ASIS is observed across the published experiments with statistical significance. It helps experimenters to limit their attention to main effects when experimental resources such as time and budget are limited. Furthermore, ASIS has a large potential to improve current DOE and RPD models therefore lead to new methods and improved analysis.

The sixth chapter draws conclusions and points out future work. A general discussion is put in here and potential research directions to follow this research are suggested. Results and impacts of this research are summarized in the major contribution section. Implications and cautions are also given in the discussion.

A collection of descriptive charts and tables are attached at the end, followed by a complete reference list.

Chapter 2

Regularities in Data from Experiments

2.1 Overview

Researchers in the sciences of complexity seek to discover regularities arising in natural, artificial, and social systems and to identify their underlying mechanisms. In engineering experiments, there are also many interesting regularities existing in subject systems. These regularities appear to arise from the interplay of the physical behavior of the systems and the knowledge of the experimenters. Analysis of these regularities should be interesting to a broad range of investigators in complex systems including engineers, statisticians, physicists, cognitive scientists, and social scientists.

In order to understand the regularities arising from engineering experimental data, we have to understand certain Design of Experiments terminology and methodology. Regularities are derived based on experimenters' experience with different experiments and systems. To master these regularities, we also carry out studies with computer simulations.

This chapter focuses on Design of Experiments theory and regularities derived from experimental data. The whole chapter is organized as follows: Section 2 presents necessary background in Design of Experiments; Section 3 describes system regularities in details; Section 4 presents system models illustrating three types of regularities and how to incorporate them into

DOE simulations; Section 5 illustrates the importance of understanding and using these regularities; Section 6 gives a summary of the whole chapter.

2.2 Design of Experiments

In an engineering design process, there are many important questions which demand experiments. Some of these questions are:

- Which design variables affect my design the most?
- How do I relate my design variables to the response of my system?
- How well does the system perform in the presence of noise?
- What is the best configuration of factor values to minimize variation in a response?

To answer these questions and help us understand subject systems, we usually plan and carry out a series of experiments and analyze their results. Those experiments take several different formats and include physical experiments, computer simulations, and experiments which combine them together.

However, experiments can be very costly and time consuming. The first line of defense against experimentation goes typically like the following: “Doing an experiment would be incredibly expensive" or “For doing this right, I would need hundreds of subjects, I would be busy for years without being able to finish one project, and the cost would be enormous." Furthermore, experimental results should be subject to an analysis in a scientific way so that we can get most information out of the results efficiently with confidence. All these requirements demand experiments to be designed well.

Design of Experiments (DOE) is an indispensable tool for experimenters. The mathematical and scientific discipline of Design of Experiments seeks to provide a theoretical basis for experimentation across many domains of inquiry. Commonly articulated goals of DOE include:

- Making scientific investigation more effective and reliable [3];
- Efficient process and product optimization [23];
- Improvement of system robustness to variable or uncertain ambient conditions, internal degradation, manufacturing, or customer use profiles [4, 24, 25].

The use of DOE in engineering appears to be rising as it is frequently disseminated through industry "Six Sigma" programs, corporate training courses, and university engineering curricula.

A clear definition of DOE comes from the Engineering Statistics Handbook: "Design of experiments (DEX or DOE) is a systematic, rigorous approach to engineering problem-solving that applies principles and techniques at the data collection stage so as to ensure the generation of valid, defensible, and supportable engineering conclusions. In addition, all of this is carried out under the constraint of a minimal expenditure of engineering runs, time, and money." [26]

2.2.1 A Historical Review

The history of DOE can be traced back to the pioneering work of R. A. Fisher [27] who first established experimentation as a rigorous subject of study motivated by the needs of efficient agricultural experimentation in 1930s. Later on, DOE got rapidly developed after World War II. In 1978, Box, Hunter, and Hunter published a quite influential book Statistics for Experimenters [3], which became the major reference work on the design of experiments for statisticians for years afterwards. Response Surface Methodology was an important practical advance which brings optimization of systems into the concept of DOE [23, 28]. Taguchi pioneered the use of robust parameter design in which systems (a system can be a products or process) are made less sensitive to noise due to manufacturing variations, customer use, and degradation over time [24, 25]. Statisticians have continued robust design in DOE [4] and DOE has become a sophisticated field and active research area today.

Some of the advances in recent theoretical DOE development are [6]:

1) Using robust parameter design to off line quality control and productivity improvement;
2) Using the minimum aberration criterion for optimal assignment of factors to columns of a design table. This criterion is more powerful than the maximum resolution criterion for choosing fractional factorial designs;
3) The increasing use of designs with complex aliasing in conducting economical experiments;
4) Widespread use of generalized linear models and Bayesian methods for analyzing non-normal data.

2.2.2 A Technical Review

To make the discussion clear, the following definitions are provided:

- **System** – A product or process which is the subject of an experiment
- **Response** – An output of the system to be measured in an experiment.
- **Factor** – A variable which is controlled by the experimenter to determine its effect on the response.
- **Active factor** – A factor that experiments reveal to have a significant effect on the system response.

- **Level** – The discrete values a factor may take in an experiment.
- **Main effect** – The individual effects of each factor in an experiment [29]. In the 2^k design[4], the main effect of a factor is computed by averaging of all the responses at each level of that factor and taking the difference.
- **Interaction** – The failure of a factor to produce the same effect at different levels of another factor [29]. An interaction that can be modeled as arising from the joint effect of two factors is called a two-factor interaction. Similarly, three-factor interactions and higher order interactions may be defined.
- **Full factorial experiment** – An experiment in which every possible combination of factor levels is tested. In a system with k factors each having two levels, the full factorial experiment is denoted as the *2^k design*.
- **Fractional factorial experiment** – An experiment in which only an adequately chosen fraction of the treatment combinations required for the full factorial experiment is selected to be run.

For example, Table 1 depicts a 2^3 factorial experiment which is used for systems with three factors each having two levels which are coded as +1 and -1. If the main effects of the control factors are additive, a full factorial experiment provides the same precision of effect estimation "as if the whole experiment were dedicated to a single factor" [27]. If the effects of the control factors are not simply additive, a full factorial design enables the experimenter to estimate all the interactions among the factors. A significant disadvantage of the full factorial design is that the number of experiments required rises geometrically with the number of factors.

[4] Please refer to "Full factorial experiment" below.

Table 1 A full factorial 2^3 design

Trial	A	B	C
1	-1	-1	-1
2	-1	-1	+1
3	-1	+1	-1
4	-1	+1	+1
5	+1	-1	-1
6	+1	-1	+1
7	+1	+1	-1
8	+1	+1	+1

In order to conduct experiments more efficiently, fractional factorial experiments may be employed which retain the balance and orthogonality of full factorial designs, but reduce the size of experiments by decreasing its ability to resolve interactions. For example, let's consider an experimental scenario for estimating main effects of seven variables each having two discrete levels using eight experiments. The 2^{7-4} design is D-optimal[5] for fitting a first-order model -- the design minimizes the volume of the ellipsoidal confidence region of the main effect estimates. In the 2^{7-4} fractional factorial design, each main effect is aliased with a two-factor interaction. This design is said to have *Resolution III* and is therefore sometimes denoted 2_{III}^{7-4}. A design constructed so that main effects are clear of two-way interactions is said to have *Resolution IV*. Higher resolution requires more experimental runs or fewer experimental factors. For example, a 2_{IV}^{4-1} can be constructed by striking out columns C, E, and G from Table 2 below. Alternatively,

[5] A D-optimal design is one that maximizes the determinant of Fisher's information matrix, $X^T X$. This matrix is proportional to the inverse of the covariance matrix of the parameters. So maximizing $det(X^T X)$ is equivalent to minimizing the determinant of the covariance of the parameters. A D-optimal design minimizes the volume of the confidence ellipsoid of the regression estimates of the linear model parameters, β.

a 2_{IV}^{7-3} can be formed by "folding over" the 2_{III}^{7-4} which involves adding eight more runs with the opposite settings.

Table 2 A fractional factorial design 2^{7-4}

Trial	A	B	C	D	E	F	G
1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	+1	+1	+1	+1
3	-1	+1	+1	-1	-1	+1	+1
4	-1	+1	+1	+1	+1	-1	-1
5	+1	-1	+1	-1	+1	-1	+1
6	+1	-1	+1	+1	-1	+1	-1
7	+1	+1	-1	-1	+1	+1	-1
8	+1	+1	-1	+1	-1	-1	+1

2.2.3 Adaptive One-Factor-at-A-Time (OFAT) Experiments

To change one factor at a time in experiments is another experimental strategy. In this strategy, experimenters change a single factor in each experimental trial to gain information about this factor. The response is optimized by study all factors in such a way in turn. The implementation of the strategy follows several steps:

- Pick up a baseline set of factor levels with experimenters' experience and analysis
- Measure the baseline response
- In sequence for each factor in turn
    - Set the factor to each of its levels and keeping all other experimental factors constant
    - Measure the response
    - Retain the factor level that provided the best response so far

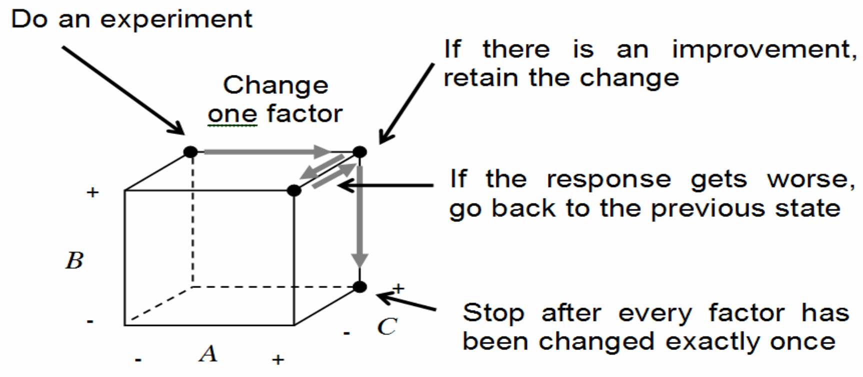


Figure 3 One-Factor-at-A-Time experiments

The whole process is illustrated as in Figure 3. This figure is adapted from Frey et al (2006) [30].

Table 3 presents an example of the adaptive one-factor-at-a-time method. A same seven-factor system as analyzed in Table 2 is studied here and the planning table can be compared to Table 2 for difference.

Table 3 A One-Factor-at-A-Time design

Trial	A	B	C	D	E	F	G
1	-1	-1	-1	-1	-1	-1	-1
2	+1	-1	-1	-1	-1	-1	-1
3	+1	+1	-1	-1	-1	-1	-1
4	-1	+1	+1	-1	-1	-1	-1
5	+1	+1	+1	+1	-1	-1	-1
6	+1	+1	+1	+1	+1	-1	-1
7	+1	+1	+1	+1	+1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1

The adaptive one-factor-at-a-time method requires n*(k-1)+1 experimental trials given n factors each having k levels. The method provides estimates of the conditional main effects of each experimental factor but cannot resolve interactions among experimental factors. The adaptive one-factor-at-a-time approach provides no guarantee of identifying the optimal factor settings. Both random experimental error and interactions among factors may lead to a sub-optimal choice of factor settings. Because of these properties, it was usually not recommended within the statistics and design methodology literature. A summary of these reasoning to discourage OFAT can be found in Frey (2003) [31]:

- It cannot estimate some interactions;
- The conclusions from its analysis are not general;
- It can miss optimal settings of factors;
- It essentially rules out the possibility of randomization and therefore can be susceptible to bias due to time trends.

However, recent research found OFAT could be more effective than fractional factorial design in certain scenario. Frey et al. (2003) [11] showed that an adaptive one-factor-at-a-time method can be effective for improving a response when pure error is not large or interactions are not small. Frey and Wang (2006) [30] proved that adaptive OFAT exploits main effects with high probability and also exploits two factor interactions if they are large. Later in this dissertation, a new adaptive approach will be incorporated into robust designs and we will show that adaptive method for robust parameter design performs better on average and more consistently than the next best alternative method, which is a cross array [2, 32].

2.3 Introduction to System Regularities

Based on experience in planning and analyzing many experiments, practitioners and researchers in DOE have identified regularities in the interrelationships among factor effects and interactions. Such regularities are frequently used to justify experimental design and analysis strategies [6]. This section reviews three regularities noted in the DOE literature, explores their nature, origins, and influence on DOE theory and practice. These three regularities are Effect Sparsity, Hierarchy, and Heredity.

2.3.1 Effect Sparsity

Effect Sparsity refers to the observation that number of relatively important effects in a factorial experiment is generally small [33]. This is sometimes called the Pareto principle in Experimental Design based on analogy with the observations of the 19th century economist Vilfredo Pareto who argued that, in all countries and times, the distribution of income and wealth follows a logarithmic pattern resulting in the concentration of resources in the hands of a small number of wealthy individuals.

Effect sparsity appears to be a phenomenon characterizing the knowledge of the experimenters more so than the physical or logical behavior of the system under investigation. Investigating an effect through experimentation requires an allocation of resources -- to resolve more effects typically requires more experiments. Therefore, effect sparsity is in some sense an indication of wasted resources. If the important factor effects could be identified during planning, then those effects might be investigated exclusively, resources might be saved, and only significant effects would be revealed in the analysis. However, experimenters are not normally able to do this. Effect sparsity is therefore usually evident, but only after the experiment is complete and the data have been analyzed.

Researchers in DOE have devised means by which the sparsity of effects principle can be exploited to seek efficiencies. Many experiments are designed to have projective properties so that when dimensions of the experimental space are collapsed, the resulting experiment will have desired properties. For example, the fractional factorial 2^{3-1} design may be used to estimate the main effects of three factors A, B, and C. As Figure 4 illustrates, if any of the three dimensions associated with the factors is collapsed, the resulting design becomes a full factorial 2^2 experiment in the two remaining factors. Projection, in effect, removes a factor from the experimental design once it is known to have an insignificant effect on the response. Projective properties of fractional factorial experiments can enable an investigator to carry out a full factorial experiment in the few critical factors in a long list of factors without knowing a priori which the critical few are out of those many factors. Similarly, Latin Hypercube Sampling enables an experimenter to sample an n-dimensional space so that, when n-1 dimensions collapse, the resulting sampling is uniform in the remaining dimension [34]. Latin Hypercube Sampling has become popular for sampling computer simulations of engineering systems suggesting that its projective properties provide substantial practical advantages for engineering design.

Although effect sparsity is widely accepted as a useful regularity, better quantification is certainly needed. Reliance on effect sparsity has led to strong claims about single array methods of robust design, but field investigation have shown that crossed arrays give better results as we previously discussed [22]. Degrees of reliance on effect sparsity may be the root cause of some disagreements about methodology in robust design.

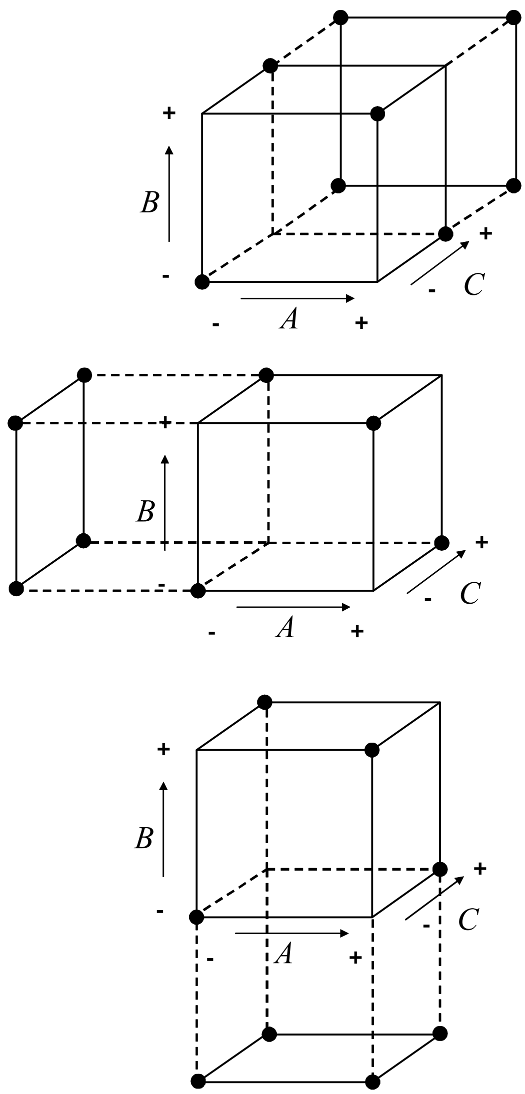


Figure 4 A fractional factorial 2^{3-1} design and its projections into 2^2 designs.

2.3.2 Effect Hierarchy

Effect Hierarchy[6] is a term denoting the observation that main effects tend to be larger on average than two-factor interactions, two-factor interactions tend to be larger on average than three-factor interactions, and so on [35]. Effect hierarchy is illustrated in Figure 5 for a sample system with four factors A, B, C, and D. Figure 5 illustrates a case in which hierarchy is not strict – for example, that some interactions (such as the two-factor interaction AC) are larger than some main effects (such as the main effect of B).

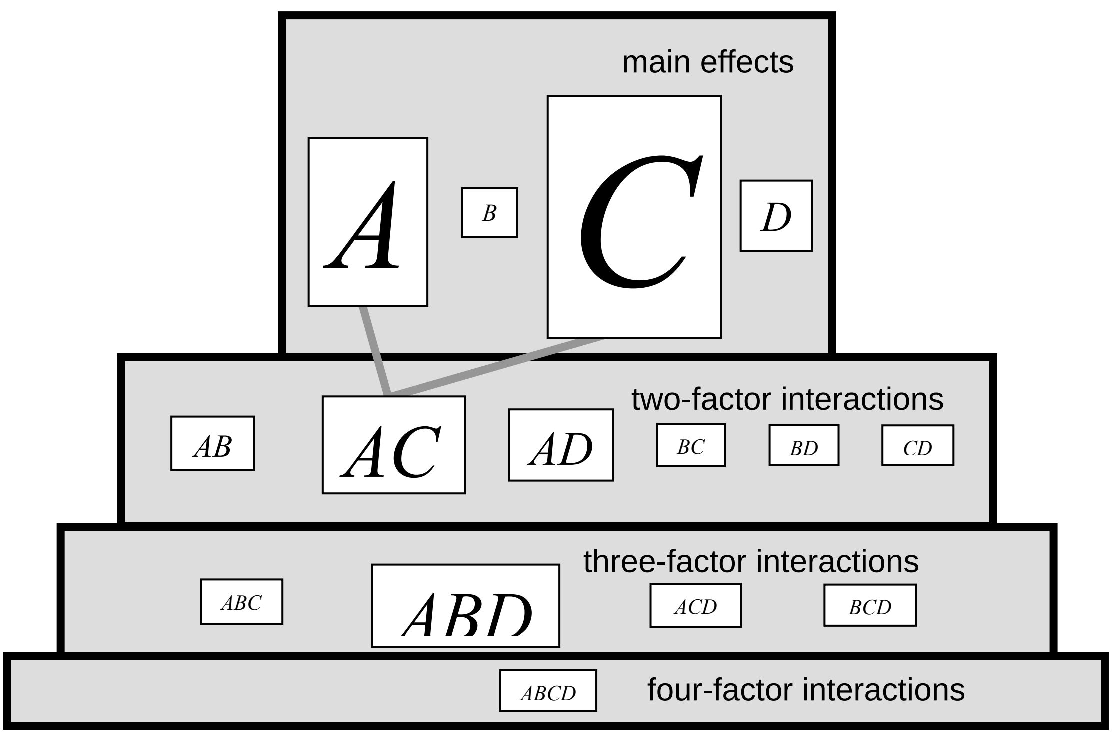


Figure 5 Effect hierarchy and heredity among main effects and interactions in a system with four factors A, B, C, and D. The font size represents the size of the effects.

[6] Effect Hierarchy is also referred as "hierarchical ordering" or simply "hierarchy".

The phenomenon of hierarchical ordering is partly due to the range over which experimenters typically explore factors. In the limit that experimenters explore small changes in factors and to the degree that systems exhibit continuity of responses and their derivatives, linear effects of factors tend to dominate. Therefore, to the extent that hierarchical ordering is common in experimentation, it is due to the fact that many experiments are conducted for the purpose of minor refinement rather than broad scale exploration.

The phenomenon of hierarchical ordering is also partly determined by the ability of experimenters to transform the inputs and outputs of the system to obtain a parsimonious description of system behavior [36]. For example, it is well known to aeronautical engineers that the lift and drag of wings is more simply described as a function of wing area and aspect ratio than by wing span and chord. Therefore, when conducting experiments to guide wing design, engineers are likely to use the product of span and chord (wing area) and the ratio of span and chord (the aspect ratio) as the independent variables. In this scenario, one might say that the experimenters have performed a non-linear transformation of input variables (span and chord) prior to conducting the experiments. In addition, after conducting the experiments, further transformations might be conducted on the response variable. In aeronautics, lift and drag are often transformed into a non-dimensional lift and drag coefficients by dividing the measured force by dynamic pressure and wing area. It is also common in statistical analysis of data to apply transformations such as a logarithm as part of exploration of the data. A key aspect of hierarchical ordering is its dependence on the perspective and knowledge of the experimenter as well as conventions in reporting data. It is important in assessing regularities in published experimental data that we do not alter the data as it was presented in any ways that affect its hierarchical structure.

Effect hierarchy has a substantial effect on the resource requirements for experimentation. A full factorial 2^k experiment allows one to estimate every possible interaction in a system with k two-level factors, but the resource requirements grow exponentially as the number of factors rises. A saturated, resolution III fractional factorial design allows one to estimate main effects in a system with k two-level factors with only $k+1$ experiments, but the analysis may be seriously compromised if there are large interaction effects in the system. Better quantification of effect hierarchy is surely needed. For example, the degree to which systems exhibit hierarchy has been shown to strongly determine the effectiveness of robust design methodologies [32]. If such decisions among robust design methods can be based on empirical studies, further efficiencies could be achieved.

2.3.3 Effect Heredity

Effect Heredity[7] implies that, in order for an interaction to be significant, at least one of its parent factors[8] should be significant [35]. This regularity can strongly influence sequential, iterative approaches to experimentation. For example, in Response Surface Methodology (RSM), high resolution experiments (e.g., central composite designs) are frequently used with a small number of factors only after screening and gradient-based search bring the response into the neighborhood where interactions among the active factors are likely. Effect heredity can also provide advantages in analyzing data from experiments with complex aliasing patterns, enabling experimenters to identify likely interactions without resorting to high resolution designs [37].

[7] Effect Heredity is also referred as “effect inheritance” or simply “inheritance”.
[8] An interaction’s parent factors refer to factors which join together to produce the interaction, e.g., factor A and C are two parent factors of the interaction AC.

The effect structures listed above have been identified through long experience by the DOE research community and by practitioners who plan, conduct, and analyze experiments. The effect structures figure prominently in discussion of DOE methods including their theoretical underpinnings and practical advice on their use. However, effect structures have not been quantified rigorously. Furthermore, there has been little effort to search for other regularities that may exist in experimental data across many domains. These gaps in the literature motivated the investigation described in the next chapter.

2.4 DOE Models Incorporating System Regularities

In the DOE simulations, statistical system models are generally used to mimic the experimental process and generate experimental responses. We introduce some models here which are widely used in DOE analysis and simulations.

2.4.1 The General Linear Model

A General Linear Model (GLM) is frequently discussed in statistical community. As shown in its name, it is a general model which represents the response of a system in a linear combination of effects of the experimental factors. In Design of Experiments, the general linear model takes a form of a polynomial. If experiments only assigned two levels to each factor, then an appropriate model should include only selected polynomial terms in Equation 1:

$$y(x_1, x_2, \dots, x_n) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j>i}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j>i}^{n} \sum_{k>j}^{n} \beta_{ijk} x_i x_j x_k + \dots + \varepsilon \quad \textbf{(1)}$$

The term β_0 is a constant which represents the mean of the response. The terms β_i represent the main effects of the factors x_i on the system response. The terms β_{ij} determine the effects of two-factor interactions with parent factors x_i and x_j. Similarly, terms β_{ijk} denote the effects of three-factor interactions. In two-level designs, the input variables are frequently normalized into coded levels of -1 and +1. Given this normalization, the sizes of the coefficients β can be compared directly to asses the relative influence of factor effects.

2.4.2 The Relaxed Weak Heredity Model

Chipman, et al. [37] proposed a model for Bayesian variable selection approach in DOE in 1997. The model can be described mathematically with Equation 2 to 9 below.

$$y(x_1, x_2, \dots, x_{14}) = \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j>i}}^{n} \beta_{ij} x_i x_j + \varepsilon \quad \textbf{(2)}$$

$$x_i \in \{+1, -1\} \;\; i \in 1 \dots n \quad \textbf{(3)}$$

$$\delta_i \in \{0,1\}; \;\; \delta_{ij} \in \{0,1\}; \;\; i, j \in 1 \dots n \quad \textbf{(4)}$$

$$\varepsilon \sim NID(0, \sigma) \quad \textbf{(5)}$$

$$f(\beta_i | \delta_i) = \begin{cases} N(0,1) & \text{if} \;\; \delta_{\mathrm{i}} = 0 \\ N(0,c) & \text{if} \;\; \delta_{\mathrm{i}} = 1 \end{cases} \quad \textbf{(6)}$$

$$f(\beta_{ij} | \delta_{ij}) = \begin{cases} N(0,1) & \text{if} \;\; \delta_{ij} = 0 \\ N(0,c) & \text{if} \;\; \delta_{ij} = 1 \end{cases} \quad \textbf{(7)}$$

$$\Pr(\delta_i = 1) = p \quad \textbf{(8)}$$

$$\Pr(\delta_{ij} = 1 | \delta_i, \delta_j) = \begin{cases} p_{00} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} = 0 \\ p_{01} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} = 1 \\ p_{11} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} = 2 \end{cases} \tag{9}$$

Equation 2 represents the relationship between system responses y and factor effects x. The independent variables x_i are factors which have potential impacts on system responses. Those factors are set particular values by experimenters and in two-level experiments there are two values set to each factor, e.g., a high temperature and a low temperature. Therefore, these two values are usually coded as +1 / -1 as desribed in Equation 3. Among those many factors, only a few of them are significant to system responses according to effect sparsity. Equation 4 shows this property. A prior δ is assigned to each effect which represents whether this effect is significant ($\delta = 1$) or not ($\delta = 0$). The variable ε represents pure experimental errors in observations of the response and it is assumed to be normally distributed as shown in Equation 5. The response is assumed to be a second order polynomial in the independent variables x_i. The coefficients β_i are the main effects and the coefficients β_{ij} model two-way interactions with their absolute value indicate how large those effects are. Although in Equation 6 and 7 those effects are models as normally distributed random variables with different variance according to their significance, these effects will be fixed once a system is generated or modeled. Equation 8 and 9 give the conditional probability of effects significance conditioned on their prior δ.

The relaxed weak heredity model was developed for the Bayesian variable selection method and reflected certain system regularities such as effect sparsity and heredity. However, it is not suitable for the robust design validation and comparison purpose. For example, it does not discriminate control factors and noise factors; the effect hierarchy property is not directly reflected in the model; and it does not contain three factor interactions which are essential under certain scenarios according to our analysis. Therefore, we developed the Hierarchical Probability Model.

2.4.3 The Hierarchical Probability Model

In order to carry out a validation and comparison on robust design methods, it is necessary to have a large number of engineering systems which are to be the subject of robust design experiments. For this purpose, each system must be able to accept as inputs the values of the control factors which represent the parameter design of the system. The control factors are assumed to take a limited set of levels thus defining a finite, but perhaps very large, set of parameter design alternatives. The model must also accept as inputs the values of the noise factors which represent the conditions in the environment, customer use, internal deterioration, and so on, to which the design is to be made robust. The model, given any specific set of control and noise factor values, must output a response value. The relationship of the response to the inputs must have properties typical of actual engineering systems such as sparsity of effects, hierarchy, and inheritance. According to these requirements, we created a model based on the weak heredity model or RWH model. Our model is expressed in Equations 10 through 19 below.

$$y(x_1, x_2, \ldots, x_{14}) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j>i}}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j>i}}^{n} \sum_{\substack{k=1 \\ k>j}}^{n} \beta_{ijk} x_i x_j x_k + \varepsilon \qquad \textbf{(10)}$$

$$x_i \sim NID(0, w_1) \;\; i \in 1 \ldots m \tag{11}$$

$$x_i \in \{+1, -1\} \;\; i \in m+1 \ldots n \tag{12}$$

$$\varepsilon \sim NID(0, w_2) \tag{13}$$

$$f(\beta_i | \delta_i) = \begin{cases} N(0,1) & \text{if} \quad \delta_{\mathrm{i}} = 0 \\ N(0,c) & \text{if} \quad \delta_{\mathrm{i}} = 1 \end{cases} \tag{14}$$

$$f(\beta_{ij} | \delta_{ij}) = \begin{cases} N(0, s_1) & \text{if} \quad \delta_{ij} = 0 \\ N(0, c \cdot s_1) & \text{if} \quad \delta_{ij} = 1 \end{cases} \tag{15}$$

$$f(\beta_{ijk} | \delta_{ijk}) = \begin{cases} N(0, s_2) & \text{if} \quad \delta_{ijk} = 0 \\ N(0, c \cdot s_2) & \text{if} \quad \delta_{ijk} = 1 \end{cases} \tag{16}$$

$$\Pr(\delta_i = 1) = p \tag{17}$$

$$\Pr(\delta_{ij} = 1 | \delta_i, \delta_j) = \begin{cases} p_{00} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} = 0 \\ p_{01} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} = 1 \\ p_{11} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} = 2 \end{cases} \tag{18}$$

$$\Pr(\delta_{ijk} = 1 | \delta_i, \delta_j, \delta_k) = \begin{cases} p_{000} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} + \delta_{\mathrm{k}} = 0 \\ p_{001} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} + \delta_{\mathrm{k}} = 1 \\ p_{011} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} + \delta_{\mathrm{k}} = 2 \\ p_{111} & \text{if} \quad \delta_{\mathrm{i}} + \delta_{\mathrm{j}} + \delta_{\mathrm{k}} = 3 \end{cases} \tag{19}$$

Equation 10 represents the measured response of the engineering system y whose mean and variance might be optimized in robust design. The independent variables x_i are both control factors and noise variables; no distinction is made in this notation except via the indices. Equation 11 shows that the first set of x variables ($x_1, x_2, \ldots x_m$) are regarded as "noise factors" and are assumed to be normally distributed. Equation 3 shows that the other x variables ($x_{m+1}, x_{m+2}, \ldots x_n$) are the "control factors" which are assumed to be two level factors. The variable ε represents the pure experimental error in the observation of the response which was assumed to be normally distributed. Since control factors are usually explored over a wide range compared to the noise factors, the parameter w_1 is included to set the ratio of the control factors range to the standard deviation of the noise factors. The parameter w_2 is included to set the ratio of the standard deviation of the pure experimental error to the standard deviation of the noise factors.

The response is assumed to be a third order polynomial in the independent variables x_i. The coefficients β_i are the main effects. The coefficients β_{ij} model two-way interactions including control by noise and noise by noise interactions. Similarly, the coefficients β_{ijk} model three-way interactions including control by control by noise and control by noise by noise interactions. The weak heredity model originally did not include three-way interaction effects, but their addition was essentially useful for our investigation so we added the three factor interaction into our model.

The values of the coefficients are determined by a random process that models the properties of effect sparsity, hierarchy, and inheritance. Equation 14 determines the probability density function for the first order coefficients. Factors can be either active or inactive depending on the value (0 or 1 respectively) of their corresponding parameters δ_i. The parameter strength of active effects is assumed to be c times that of inactive effects. Similarly, Equations 15 and 16 determine the probability density function for the second order and third order coefficients respectively. In Equations 12 and 13, the hierarchy principle is reflected in the fact that second order effects are only s_1 times as strong (on average) as first order effects (s_1<1) and third order effects are only s_2 times as strong as second order effects (s_2< s_1).

Equation 17 enforces the sparsity of effects principle. There is a probability p of any main effect being active. Equations 18 and 19 enforce inheritance. The likelihood of any second order effect being active is low if no participating factor has an active main effect and is highest if all participating factors have active main effects. Thus generally one sets $p_{11} > p_{01} > p_{00}$, and so on.

The model described in this section is certainly not the only possible way to form a set of engineering systems on which to simulate robust design methods. For example, instead of using the RWH model or Hierarchical Probability Model, one might have assembled a database of data from real engineering systems. However, for the present purposes, simulated systems provided greater flexibility. For example, in the example of Chapter 4 the Hierarchical Probability Model allowed consideration of systems with seven control factors and seven noise factors. Systems with so many factors are rarely documented in adequate detail to enable the kind of approach presented here.

2.5 Effects of System Regularities on DOE Methods

System regularities we explore here are certain properties engineering systems hold. These regularities can largely affect the efficiency and effectiveness of experiments. In this section, an example is given to show how much difference these regularities can make on comparing different DOE methods. Following that, we conclude that it is very important to study these regularities, verify and quantify them before we estimate and compare performance of experimental and robust designs.

- **Example - Comparing Orthogonal Array Method (OA), One-Factor-at-A-Time Method (OFAT) and Revised-One-Factor-at-A-Time Method (ROFAT)**

Frey, et al [11], compared the degree of improvement achieved by orthogonal array method with one-factor-at-a-time method. They showed that when experimental error was small or the interactions among control factors were large, an adaptive one-at-a-time strategy tended to perform better than orthogonal arrays. Recently, Frey, et al [30] offered a theoretical support to this conclusion. As an example, we carry out a similar comparison in this section with another new method added into the comparison, i.e., the Revised-One-Factor-at-A-Time method.

ROFAT method is a revision based on OFAT method. The difference locates in the second run of the experiments. In OFAT, the second run of the experiment is to change one factor's level and keep all the other factors unchanged. Instead, in ROFAT, the second run is to change all factors' level to check whether the response is improved or not. If the response is improved, we keep all factors' level changed; if not, we reset all factors' level back to the first run's setting. After the second run, we continue to apply the OFAT scheme to run the following experiments.

Table 4 shows an example of the ROFAT method. The experimental objective is to maximize the response. All levels are changed in the second run from the first run. Since the second run's response is much larger than the first run's, we keep all changes and then change one factor at a time in the consequence runs. In the third run, the response is further improved by changing factor G's level so we keep the change. In the fourth run, the response is not improved by changing F's level so we reset F's level back and start focusing on factor E in the next run.

Table 4 A Revised-One-Factor-at-A-Time design

Trial	A	B	C	D	E	F	G	Response
1	-1	-1	-1	-1	-1	-1	-1	5.2
2	+1	+1	+1	+1	+1	+1	+1	8.9
3	+1	+1	+1	+1	+1	+1	-1	9.3
4	+1	+1	+1	+1	+1	-1	-1	7.6
5	+1	+1	+1	+1	-1	+1	-1	5.1
6	+1	+1	+1	-1	+1	+1	-1	9.6
7	+1	+1	-1	-1	+1	+1	-1	8.7
8	+1	-1	+1	-1	+1	+1	-1	9.8

The advantage of ROFAT method is that major effects are much likely to be set on the appropriate levels after the first two runs. Then responses are refined in the consequence runs. We expect this advantage will be expressed in certain scenarios.

The comparison is carried out with computer simulations. As an example, we set seven factors in the experiments. Then based on the General Linear Model described in Equation 10 – 19, we choose to only include two factor interactions into the model and assume all higher order interactions are trivial. With these settings, we have the simplified model as below.

$$y(x_1, x_2, \ldots, x_7) = \beta_0 + \sum_{i=1}^{7} \beta_i x_i + \sum_{i=1}^{7} \sum_{\substack{j=1 \\ j>i}}^{7} \beta_{ij} x_i x_j + \varepsilon \qquad \textbf{(20)}$$

$$x_i \in \{+1, -1\} \qquad i \in 1 \ldots 7 \qquad \textbf{(21)}$$

$$\varepsilon \sim N(0, e) \qquad \textbf{(22)}$$

$$\Pr(\delta_i = 1) = p_{ME} \qquad \textbf{(23)}$$

$$\Pr(\delta_{ij} = 1 | \delta_i, \delta_j) = \begin{cases} p_{00} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} = 0 \\ p_{01} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} = 1 \\ p_{11} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} = 2 \end{cases} \qquad \textbf{(24)}$$

$$f(\beta_i|\delta_i) = \begin{cases} N(0,1) & \text{if} \quad \delta_\mathrm{i} = 0 \\ N(0,c) & \text{if} \quad \delta_\mathrm{i} = 1 \end{cases} \tag{25}$$

$$f(\beta_{ij}|\delta_{ij}) = \begin{cases} N(0,s) & \text{if} \quad \delta_{ij} = 0 \\ N(0,s \cdot c) & \text{if} \quad \delta_{ij} = 1 \end{cases} \tag{26}$$

Equation 20 includes a constant term which is negligible because a constant added to the response will not affect the comparison. The system regularities are mainly reflected from the settings of parameters p_{ME}, p_{00}, p_{01}, p_{11}, and s. The relation between model parameters and system regularities is shown in Table 5.

Table 5 The relation between model parameters and system regularities

System Regularity	Related Parameter	How is it related
Effect Sparsity	p_{ME}, p_{00}, p_{01}, p_{11}	Value of the probability
Effect Hierarchy	s	s reflects the magnitude of (AB)/A
Effect Heredity	p_{00}, p_{01}, p_{11}	The difference between them

Besides these, parameter c defines how much significant factors distinguish from not significant factors. Parameter e defines how much experimental error in the responses. These are also important model features.

In this example, we present two different settings of those parameters to be used in the comparison. They are given in Table 6. A simple analysis tells us that comparing to setting 1, setting 2 offers less sparse main effect and stronger hierarchy and heredity in systems.

Table 6 Parameters used in the DOE method comparison example

	p_{ME}	p_{00}	p_{01}	p_{11}	s	c	e
Setting 1	0.25	0.15	0.20	0.25	1.0	5.0	1.0
Setting 2	0.50	0.00	0.15	0.30	0.5	5.0	1.0

To compare three experimental design methods in a fair way, we set up criteria as below:

- Each method allows 8 trials. Therefore, for orthogonal array, it is an L8 design; for both OFAT and ROFAT, we carry out 8 runs and then stop.
- The improvement is normalized in an affine match that the largest possible improvement is set to be 1 and no improvement is set to be 0.
- For each method, the experimental error is varied from 0 to the largest value defined by parameter e in 10 even steps. In such a way, we can observe how the performance is affected by the strength of experimental error.
- For each method, we run a large number of simulations (10,000 threads) and take the mean of the improvement to compare.

Figure 6 and Figure 7 show the comparison result using parameter setting 1 and 2 in Table 6.

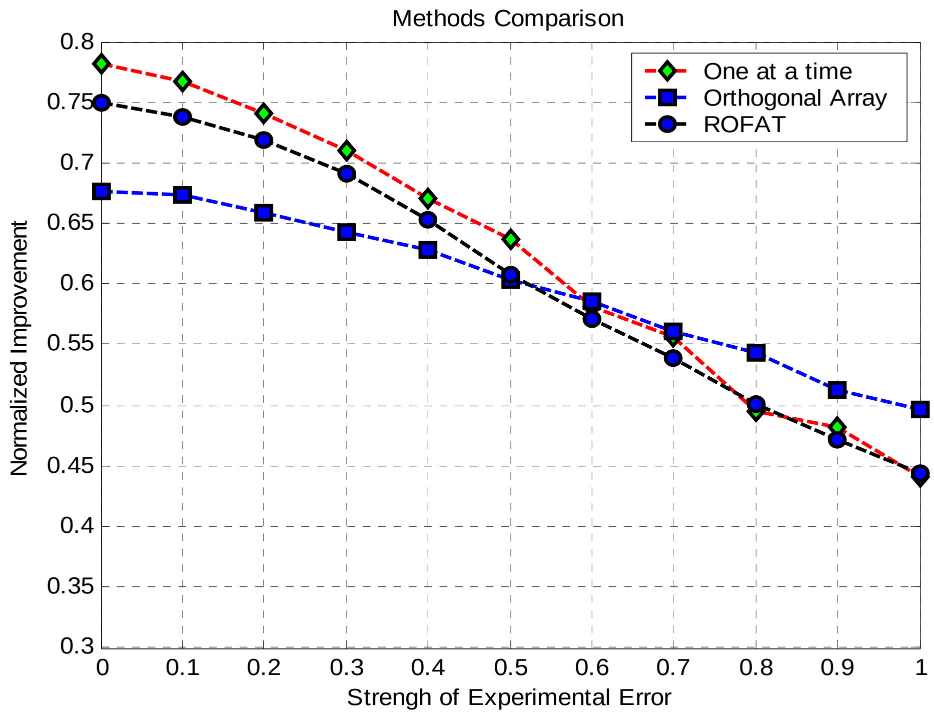


Figure 6 DOE methods comparison with parameter setting 1

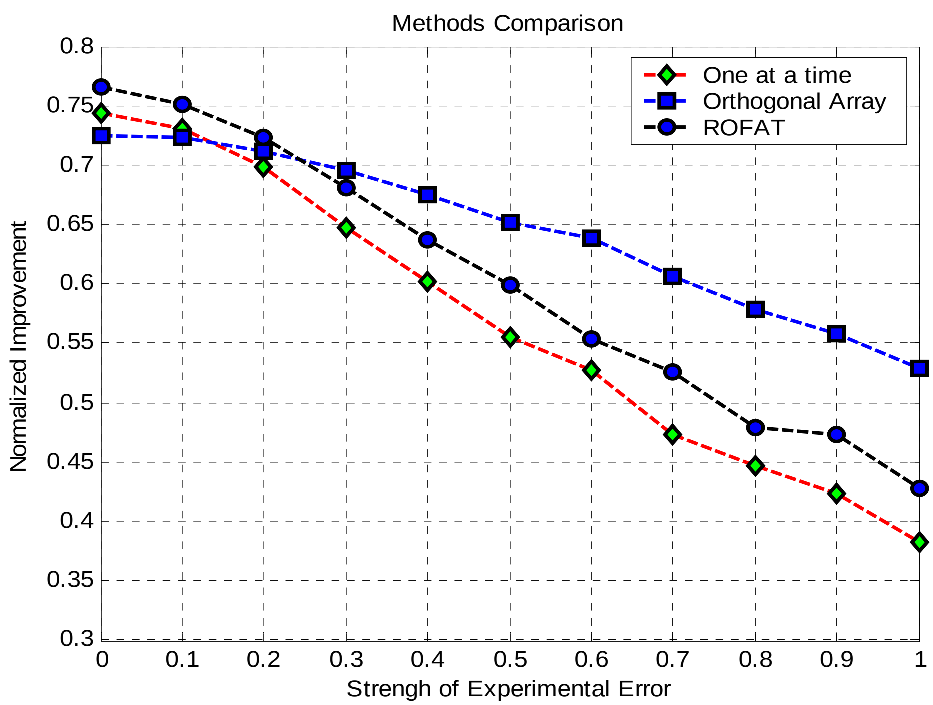


Figure 7 DOE methods comparison with parameter setting 2

From the simulation results, it shows that the strength of simulated experimental error does play a vital role in the comparison. This agrees with Frey's analysis. When the experimental error is small, OFAT performs better than OA on average. While with the experimental error increases, the performance of OFAT drops more quickly than OA and there is a crossing point which is of practical significance for experimenters.

From Figure 6 and Figure 7, we can see two obvious differences. The first one is the locations of the crossing points. With parameter setting 1, the crossing point of OFAT and OA is around 0.65 of the error strength. With parameter setting 2, whose main effects are less sparse but hierarchy and heredity are stronger, the crossing point moves ahead to about 0.12 of the error strength. And the performance gap between OFAT and OA is not as large as with setting 1 in small error strength zone. The second difference is the competence of OFAT versus ROFAT. Although ROFAT does not outperform OFAT with setting 1, ROFAT does perform constantly better than OFAT with setting 2.

The analysis shows that we have to pick up a winner method with cautious since the comparison results depend on the regularities in systems. It is almost meaningless to conclude that one method is better than another without paying attention to what system we are analyzing and what are the regularities in the system. Therefore, we will look into system regularities in the next chapter, verifying and quantifying them, searching for new regularities. All these efforts can help us understand DOE and Robust Design methods, getting sound estimations of DOE's and Robust Design's performance, and improve current methods or develop new methods.

2.6 Summary

This chapter introduces design of experiments and system regularities. An experiment is a set of actions and observations, performed in the context of solving a particular problem or question, to support or falsify a hypothesis or research concerning phenomena[9]. In practice, thousands of experiments are done in a system design process to search for optimal responses, testify functionalities, and improve system robustness. The design of experiments attempts to balance the experimental requirements and cost / time limitations so that the experiment can provide the most conclusive information with limited efforts. In design of experiments, we obey certain design criteria and apply unique methodologies such as blocking, randomization, replication, orthogonality, analysis of variance (ANOVA) and fractional factorial designs. Many of these methods are developed based on system regularities observed in experiments such as effect sparsity, hierarchy, and heredity.

System regularities are observed by practitioners empirically. However, they are not verified and quantified. An example following Frey's research on comparing different DOE methods shows that these system regularities are critical in analyzing the performance of DOE methods and comparing the effectiveness of different DOE or Robust Design methods. Different parameter values of these identified regularities bring in different system properties and affect choosing proper DOE or Robust Design methods and calculating the cost-benefit estimations. This motivates us to carry out a deep research on verifying, quantifying system regularities and searching for new regularities.

[9] Definition of "Experiments" from http://en.wikipedia.org

Chapter 3

Verification and Quantification of System Regularities

3.1 Overview

This chapter documents a meta-analysis of 113 data sets from a wide range of science and engineering disciplines. The goal is to identify and quantify regularities in the experimental data regarding the size of factor effects and interactions among factors. These regularities appear to arise from the interplay of the physical behavior of the systems and the knowledge of the experimenters. Therefore our results should be interesting to a broad range of investigators in complex systems including engineers, statisticians, physicists, cognitive scientists, and social scientists.

In this verification and quantification research on system regularities, it is shown that:

1) About 40% of main factor effects are active (have a significant influence on the system response);
2) The influence of factors separately is about five times larger on average than interactions between two factors;

3) Interactions between two factors are much more likely to be large when both participating effects are large;
4) There is a larger chance for three factor interactions to be significant than current views in the DOE literatures;
5) All three regularities are verified but they all need to be used with caution.

3.2 Objectives and Methods

The objective of this chapter is to quantify regularities observed among factor effects and multi-factor interactions. Three previously observed regularities are analyzed – effect sparsity, effect hierarchy, and effect heredity. These regularities are then incorporated into the hierarchical probability model which can be used in the study of experimental designs and be employed to simulations to testify performance of DOE and robust designs. The application details will be given in Chapter 4.

The study is performed using a set of 46 published engineering experiments which include 113 responses in all. A General Linear Model is used to estimate factor effects in each data set and the Lenth method is used to identify active effects. Then, across the set of 113 responses, the model parameters and the relevant conditional probabilities are analyzed. Details of the approach are given in the following subsections.

3.2.1 The General Linear Model Revisited

The general linear model used in this study has already been introduced in section 2.4.1, Equation 1. It takes the form as below.

$$y(x_1, x_2, \ldots, x_n) = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j>i}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j>i}^{n} \sum_{k>j}^{n} \beta_{ijk} x_i x_j x_k + \ldots + \varepsilon \qquad \textbf{(1a)}$$

The research is focusing on verifying and quantifying those coefficients and the relationship between β_i , β_{ij} , and β_{ijk} .

3.2.2 The Lenth Method for Effect Analysis

An effect in an experiment is the observed influence of a factor or combination of factors on a response. An effect is sometimes said to be "active" or "significant" if it is judged to be a significant effect by one of various proposed statistical tests. Among the commonly used test for "active" effects are the Normal Plot (or Half-Normal Plot) method [38], Box-Meyer Method [33], and the Lenth Method [39]. In this investigation, the Lenth method was selected because it is applicable to unreplicated factorial experiments, because it is computationally simple, and because it can be automated without applying many arbitrary assumptions.

In the Lenth method, a plot is made of the numerical values of all effects and a threshold for separating active and inactive effects is calculated based on the standard error of effects. In the first step, a parameter s_0 is formed

$$s_0 = 1.5 \times median\,|\,\beta\,| \tag{27}$$

where β includes all estimated effects including main effects and interactions $\beta_1, \beta_2, ..., \beta_{12}, ...$. Then the pseudo standard error (PSE) and margin of error of the effects is defined to be, respectively

$$PSE = 1.5 \times \underset{|\beta|<2.5s_0}{median}\,|\,\beta\,| \tag{28}$$

$$\text{Margin of Error} = t_{0.025,df} \times PSE \tag{29}$$

where $t_{0.025,df}$ is the 0.975th quantile of the t-distribution and df is the statistical degrees of freedom. Lenth suggests that the degrees of freedom should be one third of the total number of effects.

The margin of error for effects is defined to provide approximately 95% confidence. A more conservative measure, the simultaneous margin of error (SME) is also defined as

$$SME = t_{\gamma,df} \times PSE \tag{30}$$

where

$$\gamma = \frac{(1+0.95^{1/m})}{2}. \tag{31}$$

where m is the total number of effects.

In the Lenth Method, there are two thresholds available for discriminate active and inactive effects. They are the margin of error and the simultaneous margin of error. The simultaneous margin of error is more conservative and claims less active effects than the margin of error criteria. In this investigation, we needed to select one consistent criterion of demarcation between active and inactive effects. We judged it was more appropriate to use the margin of error as the criteria in study of full factorial experiments and that the alternative simultaneous margin of error criterion is more appropriate for screening experiments.

In the Lenth method, it is common to construct a bar graph showing all effects with reference lines at both the margin of error and at the simultaneous margin of error. An example is shown in Figure 8.

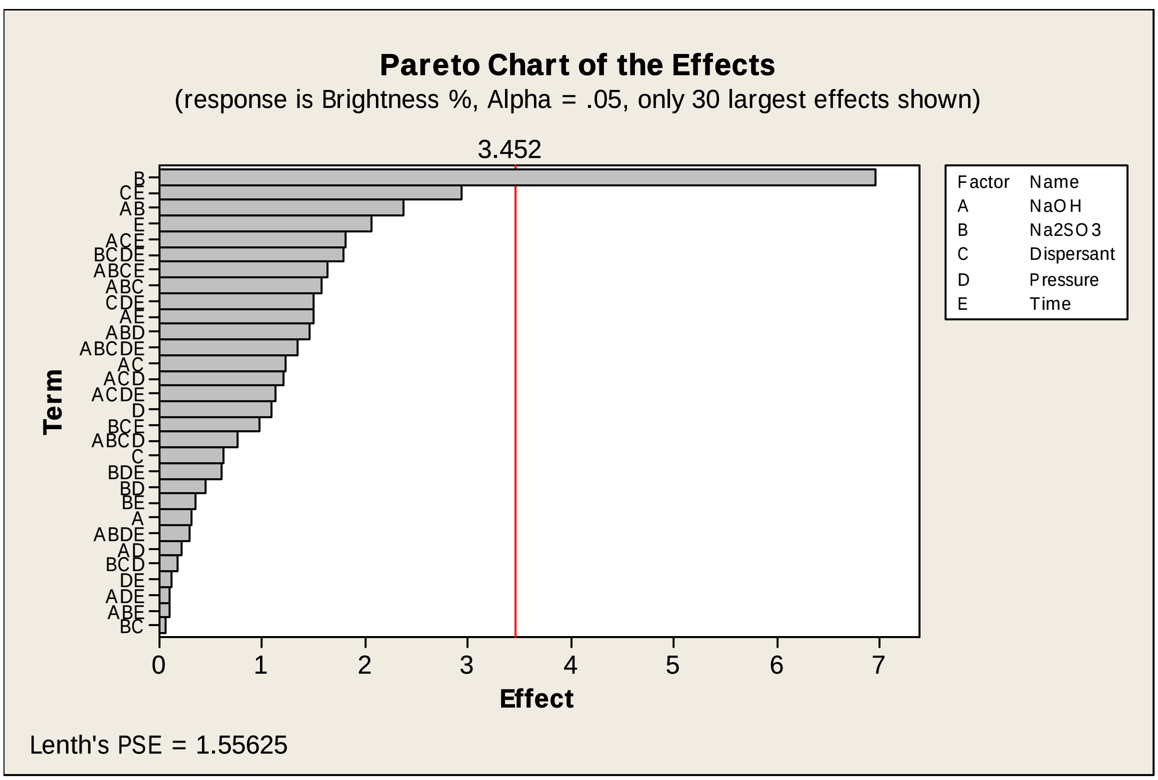


Figure 8 Bar graph of the Lenth method applied to the Laser-printed paper experiment

In the example, we use Lenth Method to analyze the data from a steam-exploded laser-printed paper experiment [40]. There are five factors in the experiment. These factors are listed in Table 7.

Table 7 Factors in the laser-printed paper experiment

Factor	**Name**
A	NaOH
B	Na_2SO_3
C	Dispersant
D	Pressure
E	Time

The experiment is a 2^5 full factorial design. There are totally 31 effects in the experiments which include 5 main effects, 10 two-factor interactions, 10 three factor interactions, 5 four-factor interactions, and 1 five factor interactions. There are several responses measured in this experiment. The response used in this example is the brightness of the paper measured in percent. Lenth Method gives a PSE value 1.55625. The red line in Figure 5 represents the threshold of active effects and inactive effects. From the analysis we can see that factor B, the existence of chemical addition Na_2SO_3, is probed to be active. A normal probability plot can help us to understand this analysis. It is shown in Figure 9.

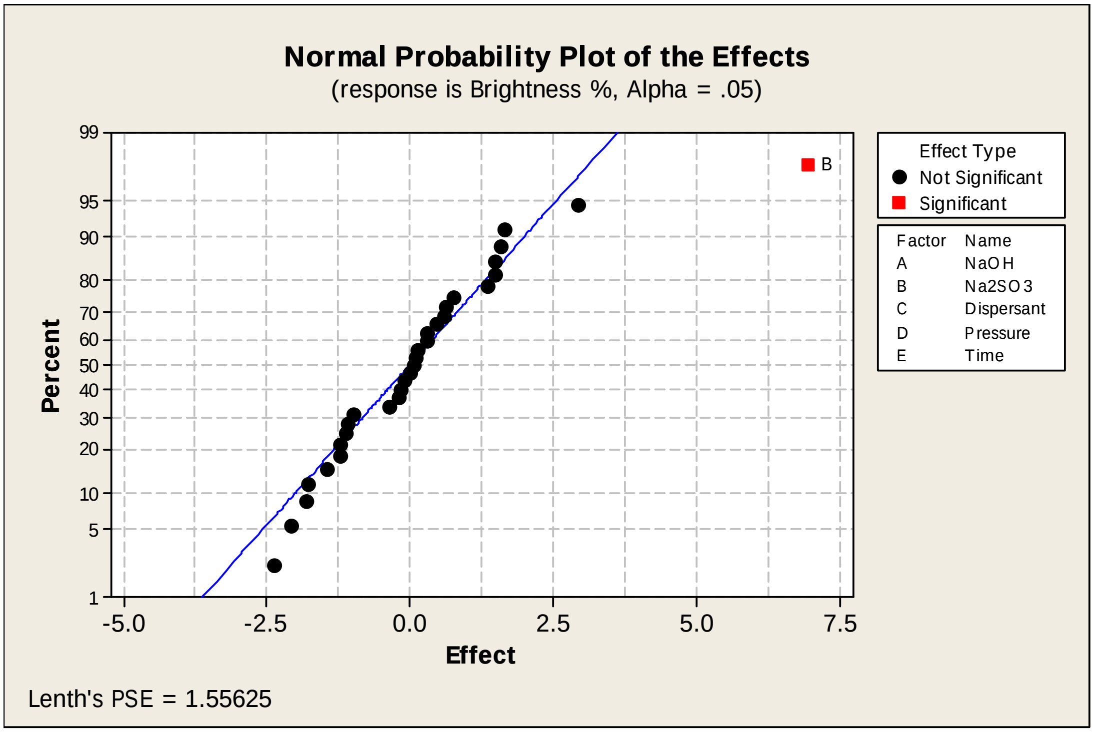


Figure 9 Normal probability plot of the Laser-printed paper experiment

3.2.3 Method for Quantifying Effect Sparsity

To quantify effect sparsity in the set of data, we used the following procedure:

1) For each experiment, estimate all the main effects and interactions as described in section 3.2.1;
2) Apply the Lenth Method and label each effect as either active or inactive as described in section 3.2.2;
3) Categorize the effects into main effects, two-factor interactions, and three-factor interactions, etc. Calculate the percentage of active effects within each category;

4) Calculate the confidence intervals (α=0.05) for the percentages of potential effects that are active. As some of the active numbers of interactions are very small, we construct exact two-sided confidence intervals based on the binomial distribution.

3.2.4 Method for Quantifying Effect Hierarchy

To test and quantify effect hierarchy, we compared the size of main effects with that of two factor interactions, and the size of two factor interactions with that of three factor interactions. As the responses in different data sets are in different units, we need to normalize them in order to make comparisons. We choose to make an affine transformation so that the minimum response and maximum response in each experiment were each, respectively, 0 and 100. This normalization was only required in our assessment of hierarchy and did not influence our assessment of other regularities discussed in this paper. The following steps summarize the procedure we used to assess hierarchy:

1) Normalize the responses of each experiment by means of an affine transformation so that they all range over the same interval [0, 100];
2) For each experiment, estimate all the main effects and interactions as described in section 3.2.1;
3) Use conventional statistical tools such as box-plots to analyze the absolute values of the main effects, two factor interactions, and three factor interactions;
4) Calculate the ratio between main effects and two factor interactions, two factor interactions and three factor interactions.

3.2.5 Method for Quantifying Effect heredity

To quantify effect heredity in the set of data we analyze probabilities and conditional probabilities of effects being active. Adopting the definitions and terminology of Chipman, et al. [37], we define p as the probability that a main effect is active, and define a set of conditional probabilities for two factor interactions

$$p_{00} = \Pr(AB \text{ is active} \mid \text{neither } A \text{ nor } B \text{ is active}) \quad \textbf{(32)}$$

$$p_{01} = \Pr(AB \text{ is active} \mid \text{either } A \text{ or } B \text{ is active}) \quad \textbf{(33)}$$

$$p_{11} = \Pr(AB \text{ is active} \mid \text{both } A \text{ and } B \text{ are active}) \quad \textbf{(34)}$$

Extending the terminology of Chipman, et. al. [37], we defined conditional probabilities for three-factor interactions as

$$p_{000} = \Pr(ABC \text{ is active} \mid \text{none of } A, B, C \text{ are active}) \quad \textbf{(35)}$$

$$p_{001} = \Pr(ABC \text{ is active} \mid \text{one of } A, B, C \text{ is active}) \quad \textbf{(36)}$$

$$p_{011} = \Pr(ABC \text{ is active} \mid \text{two of } A, B, C \text{ are active}) \quad \textbf{(37)}$$

$$p_{111} = \Pr(ABC \text{ is active} \mid \text{all of } A, B, C \text{ are active}) \quad \textbf{(38)}$$

Based on these definitions, we estimate the conditional probabilities as the frequencies observed our set of 113 responses and associated factor effects.

3.3 The Set of Experimental Data

We assembled a set of 46 full factorial 2^k experiments published in academic journals or textbooks. A list of references of all these experiments can be found in Appendix I. The experiments come from a variety of fields including biology, chemistry, materials, mechanical engineering, and manufacturing. Figure 10 uses a pie chart to present a statistics of the fields in our collected data sets. The number in the figure represents how many systems are in the according field.

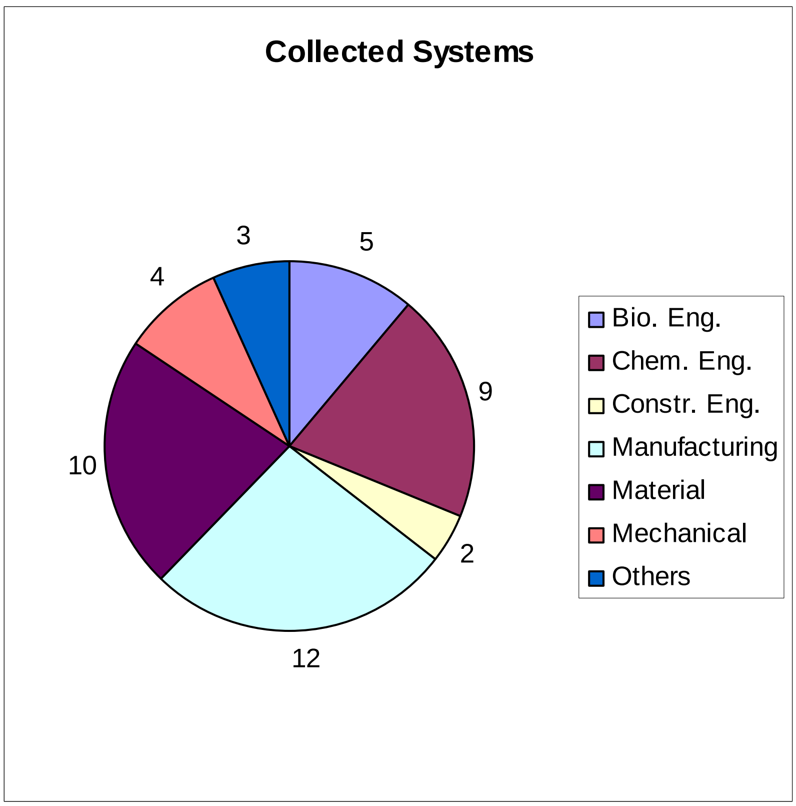


Figure 10 Statistics on fields of the engineering experiment database[10]

[10] 1) Bio. = Biological; Chem. = Chemical; Constr. = Constructional;
2) The legend labels are listed in the order that starts at 12 o'clock and to the clockwise direction on the pie chart

The reason we used full factorial designs is that we did not want to assume the existence of any given effect structure in this investigation, we want to test it and quantify it. Full factorial experiments allow all the interactions in a system to be estimated. The reason that we used two-level experiments is that they are much more common in the literature than other full factorial experiments and we wanted a large sample size.

Many of the 46 experiments contain several different responses since a single set of treatments may affect many different observable variables. Our set of 46 experiments includes 113 responses in all. Appendix II provides a complete list of these responses. Table 8 summarizes some relevant facts about the overall set. For example, Table 8 reveals that the vast majority of the experiments had either 3 or 4 factors. The number of main effects and interactions are also listed, but this is not based on analysis of the data, but only on the number of effects resolved by the experimental design. It is notable that the data set includes 569 two-factor interactions and only 383 three-factor interactions because the 54 responses from 23 designs each contribute only one potential three factor interaction. Note that the one response from a 27 experiment contributes 35 potential three-factor interactions which are about 9% of the potential three-factor interactions in the entire set.

All the experimental data in this research were recorded in our database in the form they were originally reported in the literature. No non-linear transformations were preformed before entry into the database nor were non-linear transformations conducted during the meta-analysis. Therefore, the regularities we report in the result section 3.5 are regularities in data as they are presented by experimenters. They reflect properties of general engineering systems. As it is widely known in the statistics community, non-linear transformation of the response can sometimes lead to more parsimonious models and reduce active interactions. Therefore, to explore how non-linear transformations affect regularities, we conducted a follow-up study using the same methods, but performing the analysis of the data after a log transform was applied (these results are in section 3.6). This issue of transformation of data is also briefly explored via an example in Section 3.4.

Table 8 Summary of the set of 113 responses and the potential effects therein

Factors	Experiments	Responses	Potential Main Effects	Potential Two-Factor Interactions	Potential Three-Factor Interactions
3	20 (43%)	54 (48%)	162 (40%)	162 (28%)	54 (14%)
4	22 (49%)	51 (45%)	204 (49%)	306 (54%)	204 (54%)
5	2 (4%)	5 (4%)	25 (6%)	50 (9%)	50 (13%)
6	1 (2%)	2 (2%)	12 (3%)	30 (5%)	40 (10%)
7	1 (2%)	1 (1%)	7 (2%)	21 (4%)	35 (9%)
Total	46 (100%)	113 (100%)	410 (100%)	569 (100%)	383 (100%)

In quantifying the regularity of effect hierarchy, we need to compare the size of main effects with the size of interactions. Therefore, we do an affine transformation to normalize data from different experiments so that we can carry out the comparison from the same baseline. The affine transformation is completely linear. Theoretical analysis and simulations both confirm that it does not affect the effect hierarchy and other system regularities. In the affine transformation, the responses from different experiments are normalized as below.

$$\textit{Normalized}\ \text{Response} = \frac{\text{Original Response - Minimum Response}}{\text{Maximum Response - Minimum Response}} \times 100 \qquad \textbf{(39)}$$

All the responses are translated to [0 100] so that the effects derived from them can be compared with each other across experiments.

3.4 An Illustrative Example for a Single Data Set

Before presenting the meta-analysis of the complete database of 113 responses, it is helpful to observe how the method discussed in section 3 reveals the effect structures evident in a single data set. Lloyd [41] published a full factorial 2^7 experiment regarding drag torque in disengaged wet clutches. A wet clutch, such as the one depicted in Figure 11, is a device designed to transmit torque from an input shaft which is normally connected to a motor or engine to an output (which in Figure 11 is connected to the outer case). When a wet clutch pack is disengaged, it should transmit no torque and thereby create no load on the motor. In practice, wet clutch packs result in a non-zero drag torque resulting in power losses. In Figure 11, the upper half shows an application of wet clutch to motor cycles. The lower half shows a simplified illustration of how a wet clutch is designed to work.

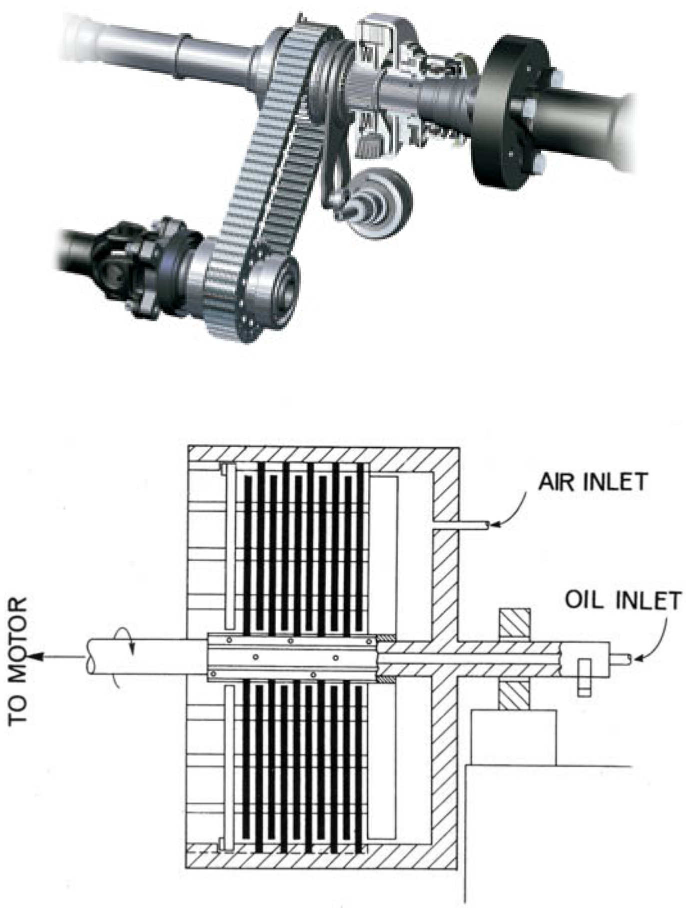


Figure 11 A wet clutch pack[11]

The study by Lloyd was conducted at *Raybestos Manhattan Inc*., a designer and manufacturer of clutches and clutch materials. The experiment was designed to assess the influence of various factors on power loss and was likely a part of a long-term effort to make

---

[11] The below figure is adapted from Lloyd [41]

improvements in the design of clutches. The factors in the study were oil flow (A), pack clearance (B), spacer plate flatness (C), friction material grooving (D), oil viscosity (E), friction material (F), and rotation speed (G). Most of these factors are normally under the control of the designer, however some of these variables such as oil viscosity might vary substantially during operation and therefore were probably included in the study to assess there influence as noise factors. However, for the purpose of the experiment, it must have been the case that all these factors were brought under the control of the experimenter to a substantial degree. Each factor was varied between two levels and the drag torque was measured as the response.

We apply the Lenth method in the analysis and show the results in Figure 12. The effects are set along the x-axis from main effects to interactions. There are dotted lines to separate the effects into seven catalogs. From left to right, they are main effects (there are 7 of them), two-factor interactions (21), three-factor interactions (35), four-factor interactions (35), five-factor interactions (21), six-factor interactions (7), and seven-factor interactions (1). The y-axis represents the size of the effects. Two bold solid lines represent the margin of error threshold. Any effect whose size is out of the band between the margin of error lines is considered to be active effect. The thin solid lines represent the simultaneous margin of error which serves as a conservative reference. Each "+" sign represents an effect in the experiment.

The complete results of the full factorial experiment are too lengthy to present here, but the main effects and active two factor interactions as determined by the Lenth method are presented in Table 9 and Table 10. This is slightly different from Lloyd's analysis in the original paper because there he simply assumed effects of order 4 or higher were all insignificant.

Figure 12 Effect analysis using the Lenth method for the wet-clutch experiment

Table 9 The main effects from the wet clutch case study.

Effect	Drag torque [ft lbs]	Active?
A	1.33	Yes
B	-1.55	Yes
C	-1.81	Yes
D	0.067	No
E	2.81	Yes
F	-0.092	No
G	3.01	Yes

Table 10 The active two-factor interactions from the wet clutch case study.

Effect	Drag torque [ft lbs]
AD	0.530
AG	0.964
BD	-0.520
BG	-0.830
CD	0.683
CG	-0.695
DE	0.642
DG	-0.914
EG	1.31

Every major effect structure under investigation in this study can be observed in this data set:

- Effect sparsity is indicated in the sense that there are 127 effects estimable within this experiment but only 21 were active which include 5 main effects, 9 two-factor interactions, and 7 higher order interactions. Effect sparsity is only weakly indicated by the main effects since 5 out of 7 were active in the study, but is strongly indicated among interactions since only 14 of 122 possible interactions were active.

- Effect hierarchy is strongly indicated because the proportion of potential effects that actually prove to be active is strongly a function of the number of factors involved. Among main effects, 5 out of 7 are active. Among two-factor interactions, 9 out of 21 are active. Among three-factor interactions, only 7 out of 35 are active.

- Effect inheritance is strongly indicated. The four largest two-factor interactions involved two factors both with active main effects. Of the remaining five two-factor interactions, all involved at least one active main effect.

Non-linear transformations of responses can strongly affect regularities in data. To illustrate this, we applied a log transformation to the drag torque of the wet clutch pack and repeated our analysis of the data. The main effects and active two factor interactions as determined by the Lenth method are presented in Tables 11 and Table 12.

Table 11 The Main Effects from the Clutch Case Study Using a Log Transform.

Effect	Log(Drag torque)	Active?
A	0.269	Yes
B	-0.350	Yes
C	-0.369	Yes
D	0.040	No
E	0.613	Yes
F	-0.015	No
G	0.529	Yes

Table 12 The active 2fi's[12] from the wet clutch case study with a log transformation.

Effect	Log(Drag torque)
AD	0.094
AG	0.159
BC	-0.072
BD	-0.096
BE	0.108
BG	-0.143
CD	0.182
CE	0.071
CF	-0.063
DE	0.103
DG	-0.228
EG	0.167

For this particular data set, the log transform failed to improve the hierarchical ordering of the data. The active main effects are not affected by the transformation. However, the number of active two factor interactions actually increased from 9 to 12. This shows that transformation affected effect sparsity and heredity in this system.

It is also important to note here that data transformation will also affect the synergistic relations between main effects and two factor interactions. This will be present later in Chapter 5.

3.5 Results of Meta-Analysis of 133 Data Sets

The methods described in section 3.2 were applied to the set of 113 responses from published experiments (Appendix I and II). Some of the main results of this meta-analysis are summarized in Table 13.

[12] We use "2fi" to refer to two-factor interactions and "3fi" to refer to three-factor interactions.

Table 13 Percentage of potential effects in 113 experiments those were active as determined by the Lenth method.

	Main effects	Two-factor interactions	Three-factor interactions	Four-factor interactions
Number of effects	410	569	383	141
Number of active effects	170	63	26	4
Percentage of effects that were active	41%	11%	6.8%	2.8%
Confidence intervals (α=0.05) on the percentage of effects that were active	37% to 46%	9% to 14%	4.5% to 9.8%	0.8% to 7.1%

The main effects were not very sparse with more than one third of main effects classified as active. However, only about 7.4% of all possible two-factor interactions were active. The percentage drops steadily as the number of factors participating in the interactions rise. Thus, Table 13 tends to validate both the effect sparsity principle (especially as applied to interactions) and also tends to validate the hierarchical ordering principle. However, this study also supports a caution in applying effect sparsity and hierarchy. For example, if about 2.2% of three-factor interactions are active (as Table 13 indicates), then most experiments with seven factors will contain one or more active three-factor interactions.

Figure 13 depicts a box plot of the absolute values of factor effects for each of three categories – main effects, two-factor interactions, and three-factor interactions. The median of main effect strength is about four times larger than the median strength of two-factor interactions. The median strength of two-factor interactions is more than two times larger than the median strength of three-factor interactions. However, Figure 13 also reveals that many two-factor and three-factor interactions were observed that were larger than the median main effect. Again, the trends in this study support the principle of hierarchy, but suggest caution in its application.

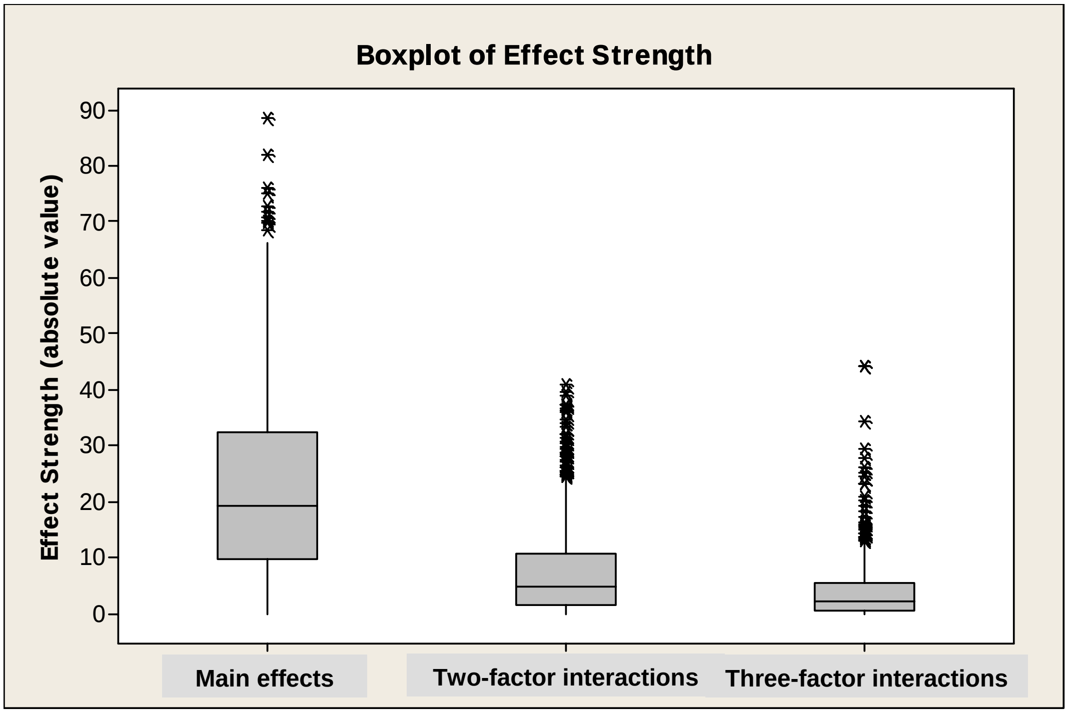


Figure 13 Box plot of absolute values for main effects, two-factor interactions, and three-factor interactions.

Table 14 presents the conditional probabilities of observing active effects. This data strongly supports the effect heredity principle. Whether the factors participating in an interaction have active main effects strongly determines the likelihood of an active interaction effect. It is noteworthy that, under some conditions, a two-factor interaction is about as likely to be active as a main effect. In addition, it is observed that, under the right conditions, a three-factor interaction can be fairly likely to be active, but still only half as likely as a main effect.

Table 14 The conditional probabilities of observing active effects based on meta-analysis of 113 experiments.

p	p_{11}	p_{01}	p_{00}	p_{111}	p_{011}	p_{001}	p_{000}
41%	33%	4.5%	0.48%	15%	6.7%	3.5%	1.2%

3.6 Quantification of the Standard Deviation *c*

In the Hierarchical Probability Model, parameter c is a very important measure which discriminates active effects from inactive effects. The model assumes inactive effect to be generated from a normal process with standard deviation 1 while active effects are generated from a normal distribution with standard deviation c. From the definition we know that $c > 1$. In this section, we derive a close form equation to estimate c based on the hierarchical probability model described in section 2.4.3 and quantify c using data sets from the engineering experiment database.

Firstly, let's define two new variables following independent normal distribution

$$R_1 \sim N(0, s^2) \tag{40}$$

$$R_2 \sim N(0, c^2 s^2) \tag{41}$$

The prior variable δ_i follows a Bernoulli's distribution with probability p, and it is independent of R_1 and R_2

$$\delta_i \sim Bernoulli(p) \tag{42}$$

Therefore, the main effect coefficients can be expressed as

$$\beta_i = (1-\delta_i)R_1 + \delta_i R_2 \quad \textbf{(43)}$$

The estimated value and variance can be directly calculated as

$$E[\beta_i] = E[(1-\delta_i)R_1 + \delta_i R_2] = E[(1-\delta_i)]E[R_1] + E[\delta_i]E[R_2] = 0 \quad \textbf{(44)}$$

$$Var[\beta_i] = E\{[(1-\delta_i)R_1 + \delta_i R_2]^2\} = (1-p+pc^2)\cdot s^2 \quad \textbf{(45)}$$

Similar processes can be applied to coefficients of 2fi's and 3fi's. For 2fi's, the variance is calculated separately according to their parent factors

$$E[\beta_{ij}] = 0 \quad \textbf{(46)}$$

$$Var[\beta_{ij} \mid \delta_i + \delta_j = 0] = [(1-p_{00}) + p_{00}c^2]\cdot s^2 \quad \textbf{(47)}$$

$$Var[\beta_{ij} \mid \delta_i + \delta_j = 1] = [(1-p_{01}) + p_{01}c^2]\cdot s^2 \quad \textbf{(48)}$$

$$Var[\beta_{ij} \mid \delta_i + \delta_j = 2] = [(1-p_{11}) + p_{11}c^2]\cdot s^2 \quad \textbf{(49)}$$

For 3fi's, the variance is also calculated in each active state of their parent factors

$$E[\beta_{ijk}] = 0 \quad \textbf{(50)}$$

$$Var[\beta_{ijk} \mid \delta_i + \delta_j + \delta_k = 0] = [(1-p_{000}) + p_{000}c^2]\cdot s^2 \quad \textbf{(51)}$$

$$Var[\beta_{ijk} \mid \delta_i + \delta_j + \delta_k = 1] = [(1-p_{001}) + p_{001}c^2]\cdot s^2 \quad \textbf{(52)}$$

$$Var[\beta_{ijk} \mid \delta_i + \delta_j + \delta_k = 2] = [(1-p_{011}) + p_{011}c^2]\cdot s^2 \quad \textbf{(53)}$$

$$Var[\beta_{ijk} \mid \delta_i + \delta_j + \delta_k = 3] = [(1-p_{111}) + p_{111}c^2]\cdot s^2 \quad \textbf{(54)}$$

All the active probabilities have been achieved in previous calculation. By calculating the variance in each catalog, we will get a group of linear equations of c^2s^2 and s^2 as Equation 55. The parameter c can be found by applying least square regression on Equations 55. The regression result is shown in Equation 56.

$$\begin{pmatrix} 866.48 \\ 235.12 \\ 45.98 \\ 108.19 \\ 191.59 \\ 26.98 \\ 16.99 \\ 13.24 \end{pmatrix} = \begin{pmatrix} 0.3863 & 0.6137 \\ 0.0048 & 0.9952 \\ 0.0450 & 0.9550 \\ 0.3300 & 0.6700 \\ 0.0118 & 0.9882 \\ 0.0353 & 0.9647 \\ 0.0667 & 0.9333 \\ 0.1510 & 0.8490 \end{pmatrix} \cdot \begin{pmatrix} c^2 s^2 \\ s^2 \end{pmatrix} \tag{55}$$

$$s^2c^2 = 1227.4; \quad s^2 = 34.3; \quad c = 5.980 \tag{56}$$

3.7 Conclusions and Discussion

Despite the possibility of exciting consequences, the results presented here must be interpreted carefully. The investigation is entirely based on full factorial experiments published in journals and textbooks. It seems plausible that biases may have influenced the data set. Full factorial experiments are most likely to be conducted for systems that have already been investigated using less resource intensive means. For example, it is common practice to use a screening experiment prior to using a higher resolution design. A specific consequence is that all the estimates of percentages of active effects in Table 14 may be inflated. If the screening stage has filtered out several inactive factors, then the experiments with the remaining factors are more likely to exhibit active effects of all kinds. In order to characterize the structure of a larger population of systems on which experiments have been conducted, responses could be selected at random from many engineering domains, and then full factorial experiments might be carried out specifically for the purpose of an extended study of system regularities and analyzed using the methods described here. Such an effort would be resource intensive, but it would guard against potential biases introduced by studying only those systems on which full factorial experiments have already been conducted.

One major outcome of this work is concerns validation and quantification of previously known regularities. All three regularities commonly discussed in the Design of Experiments literature (effect sparsity, hierarchy and heredity) were confirmed as statistically significant. In addition, each of these regularities was quantified rigorously potentially providing great practical benefits. For example, many investigators will find that, according to this study, these regularities are not as strong as they previously supposed. Although effect sparsity and hierarchy are shown to be significant trends, the exceptions to these trends are not unlikely, especially given the large number of opportunities for such exceptions in complex systems. The data presented here suggest that a system with four factors is more likely than not to contain a significant interaction given that $11\%\binom{4}{2} + 6.8\%\binom{4}{3} > 50\%$. The data also suggest that a system with a dozen factors is likely to contain around 22 active interactions with roughly double numbers of active three-factor interactions to active two-factor interactions since $6.8\%\binom{12}{3} \approx 2 \times 7 \approx 2 \times 11\%\binom{12}{2}$. This implies critical importance of three factor interactions if a system contains a large number of factors.

These observations may be important in robust parameter design. It is known that robust design relies on the existence of some two-factor interactions for its effectiveness. However, some three-factor interactions may interfere with robust design, depending on which method is used. For example, field comparisons of single array methods and crossed array methods have revealed that crossed arrays are more effective. This has led to the conjecture that single arrays rely too strongly on effect sparsity [42]. The meta-analysis in this paper suggests that the problem may be more closely related to effect hierarchy. Depending on the number of factors, three-factor interactions may be more numerous than two-factor interactions. Any robust design method that relies on strong assumptions of effect hierarchy is likely to give disappointing results unless some effective steps are taken to reduce the likelihood of these interactions through system design, response definition, or factor transformations.

Another benefit may arise from this study because it quantifies effect heredity. Bayesian methods have been proposed for analyzing data from the experiments with complex aliasing patterns [35]. These methods require prior probabilities for the parameters given in Table 14 (p_{11}, p_{01}, and so on). A hypothesis for future investigation is that using the results in Table 14 in concern with the Bayesian methods will provide more accurate system models than the same methods using previously published parameter estimates.

One of the important objectives of this investigation is to search for new regularities in engineering systems. Using our data sets, we found a statistically significant phenomenon which we described as asymmetric synergistic interaction structure (ASIS). We will fully discuss it in Chapter 5.

3.8 Summary

This chapter presented an investigation to verify and quantify the generally observed concepts of factor sparsity, hierarchy, and heredity. We built a compilation of a large number of 2^k-Designs where k is at least 3. With the data sets in this engineering experiment database, we analyzed each experiment and estimated its complete set of main effects and interactions. The Lenth method was employed to set threshold between active and inactive effects. Our results strongly support the three regularities of effect sparsity, effect hierarchy, and effect heredity. But the analysis also brought certain cautions to practitioners when their designs largely depended on these regularities. Our results can be used to refine a system model for experimental design, help in efficient planning and analysis of experiments, be integrated into computer simulations for performance estimation or method comparison, and derive informative prior values in the Bayesian model selection for analyzing experiments with aliasing patterns.

Chapter 4

Model-Based Validation and Comparison of Robust Parameter Design Methods

4.1 Overview

Uncertainty, robustness, and flexibility are three of the issues widely viewed as central to the field of complex engineered systems. A key challenge is to make the consideration of these issues more quantitative and theoretically sound. This chapter presents a detailed technical analysis of one question inter-relating all three issues. How can we quantify the benefits of flexibility in the design processes used to seek robustness to uncertainty? To address this question, an approach is presented for evaluating robust design methods via probabilistic simulation. Different robust design methods are repeated on a large number of systems and the results are analyzed statistically.

A key to the approach is appropriate modeling of interactions among the variables in complex engineering systems. The system regularities we studied in previous chapters are incorporated into probability models to generate to generate simulated systems with the properties of effect sparsity, hierarchy, and inheritance. This is a direct application of the system regularity research presented in previous chapters. At the same time, it is also a new approach to solve the difficulties in pair-comparison of robust design methods.

Two case studies are presented in which different combinations of factorial designs and adaptive plans are used in robust design experiments. It will show that a combination of an adaptive inner array (a flexible plan) with a fractional factorial outer array (an efficient but inflexible plan) provides the best results. Thus, the case studies provides specific operational advice for deploying flexibility in one particular systems engineering task. Because of this bit of progress, the new evaluation technique is proposed one example of rigorous quantitative treatment of foundational issues in complex engineering systems made possible by combining advances in computing power and statistical theory.

4.2 The Concept of Validation

It is important to clarify the concept of validation especially as it applies in engineering design. In engineering, the term "validation" is applied in different ways to models and to methods. Both uses of the term will be explored in this section.

Engineers frequently seek to evaluate a model for a specified use or a range of uses. The American Institute of Aeronautics and Astronautics (AIAA) defines model validation as "the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model" [43]. Hazelrigg proposed an alternative definition of model validity from the perspective of Decision Based Design [44]. According to Hazelrigg, a model is valid to the extent that it supports the conclusion "Design point O will produce an outcome that is preferred to the outcome that would be produced by design point C with essentially probability 1." This definition of validity emphasizes resolution versus accuracy. A desirable property of Hazelrigg's definition is that a relatively inaccurate model may be viewed as valid for making choices among alternatives when one of the alternatives is vastly preferred to the others. However, a potential drawback of Hazelrigg's definition arises from its explicit subjectivity. Hazelrigg states that "model validation is a down and dirty, personal thing" and that "a model is valid when, in the mind of the decision maker, it's up to the task." If these two quotes are taken literally, then a model is valid even if it fails to represent significant physical effects so long as the engineer is unaware of these neglected effects so that, in his mind, the model is up to the task. The AIAA definition has very different consequences with regard to the engineer's state of knowledge. The AIAA definition emphasizes the objective correspondence of the model with the real world, not the subjective belief that the model is up to the task. At the same time, the AIAA definition preserves the desirable property that a relatively inaccurate model may be viewed as valid under some conditions. The degree of correspondence with reality needed depends on "the perspective of the intended uses of the model." Thus, the AIAA definition of model validation avoids subjectivism while acknowledging the need for context.

Engineers frequently seek to evaluate a design method or software tool for a specified use or range of uses. The Institute of Electrical and Electronics Engineers (IEEE) defines validation as "confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled" [45]. Olewnik and Lewis propose an alternative definition of validation from the perspective of Decision Based Design [46]. Their definition is that for a decision support method to be valid, it must:

1) Be logical;

2) Use meaningful, reliable information;

3) Not bias the designer.

A desirable property of this definition is that it reveals the ways that some design methods impose preferences on the designer. However, a potential drawback of Olewnik and Lewis' definition is that, under this definition, determining that methods are invalid "does not imply that the methods are ineffective" [46]. By contrast, the IEEE definition emphasizes a link between validation and an assurance of effectiveness for its specific intended uses.

We propose that there is practical value in the pursuit of objective validation techniques which provide some assurance of effectiveness in application. For this reason, this paper will adopt the definitions of validation put forth by AIAA and IEEE. Although objective validation is to be pursued, we acknowledge that no design method can be expected to guarantee a particular benefit in every single project to which it is applied since there are many factors that affect the success of design projects. However, a link between a design method and its attendant benefits might be established statistically.

4.3 Robust Design Method and Methodology Evaluation

Uncertainty is an increasingly important consideration in the design of large-scale engineering systems. As systems include more components, interconnections, and variables, the effects of uncertainties (unless carefully managed) tend to accumulate and may lead to unacceptable risks. One effective countermeasure is to reduce the sensitivity of the design to uncertain or randomly varying factors (a.k.a. noise factors).

Robust Parameter Design (RPD) is a set of engineering and statistical methods for improving quality. Its purpose is to reduce the variability in performance of products and processes in the face of uncontrollable variations in the environment, manufacture, internal degradation, and usage conditions. Figure 14 illustrates a general design model. The vector $\mathbf{x} = (x_1, ..., x_n)$ is the design parameter (or the process variable) vector of the design; the vector $\mathbf{z} = (z_1, ..., z_n)$ represents the noise factors which are uncontrollable under user conditions. The output vector $\mathbf{y} = (y_1, ..., y_n)$ represents the system performance. The design objective is to set control factors ***x*** at proper values so that the whole system would be less sensitive to those noise factors ***z*** and less sensitive to the variation of design variables **x**, i.e., the system would be more "robust". To reduce system response variations by changing design parameter values has been illustrated in Figure 15.

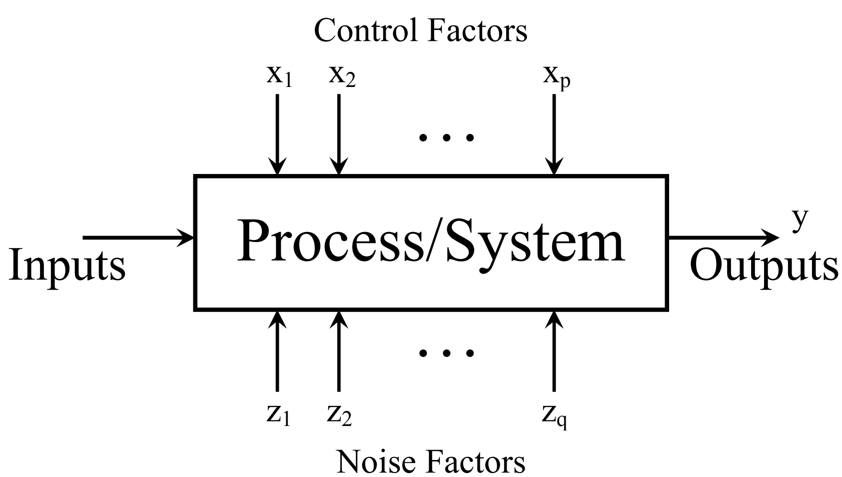


Figure 14 A general model for robust design

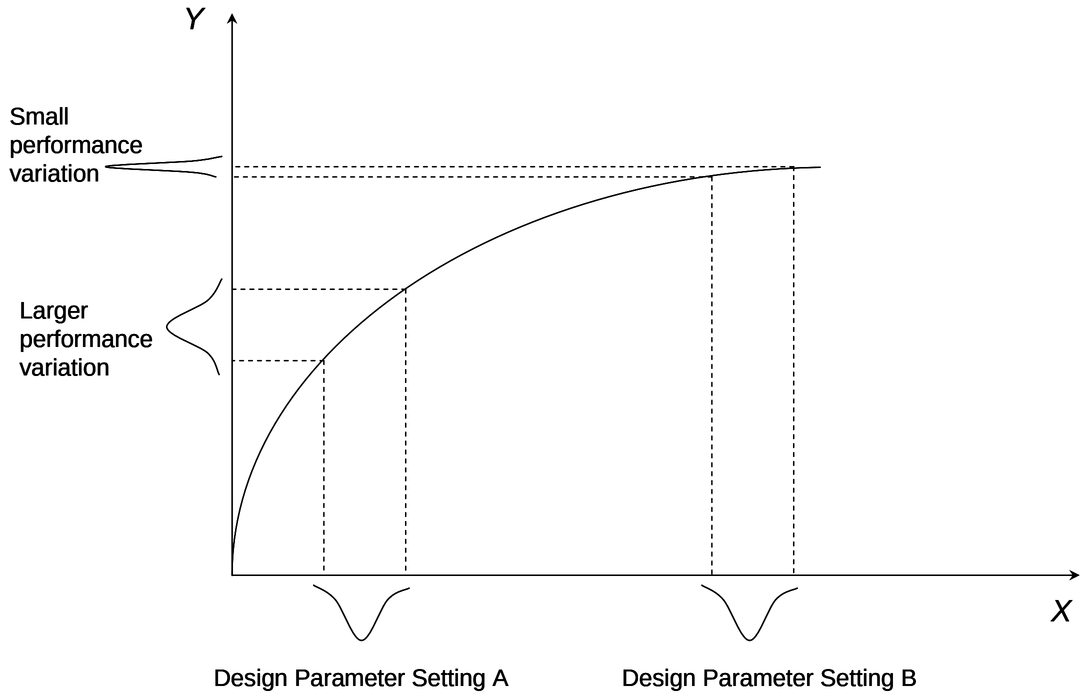


Figure 15 Adjusting design variables to reduce response variance[13]

Over the past few decades, RPD has developed through scholarly research and publication and has been widely adopted by industry. Both the industry implementation and methodological developments have been marked by considerable controversy. There is no clear consensus as to which currently available methods are most effective. It is proposed here that new ways to evaluate RPD methods can clarify many contentious issues and provide numerous insights useful for development and deployment of RPD.

Methods for robust parameter design were pioneered by Taguchi [12] through work primarily with industry in Japan. Shortly after their introduction to North America and Europe, Taguchi methods became a subject of scholarly scrutiny. As examples, we list Kacker [13],

[13] Adapted from Yang, K. and EI-Haik, B., Design for Six Sigma: A Roadmap for Product Development, McGraw-Hill, 2003, New York.

Hunter [14], Box [15, 16], Hamada and Wu [17], and Nair [18]. It was generally acknowledged that benefits emerged from Taguchi's work, but it was also widely believed that the methods could be improved through application of statistical theory. Box [15] summarized a goal of the quality engineering community to "absorb the ideas that are good" and "fix up, modify, and improve what is not so good" and "make use of the good statistical ideas developed in the West." Among the methodological prescriptions generally viewed as "not so good" were:

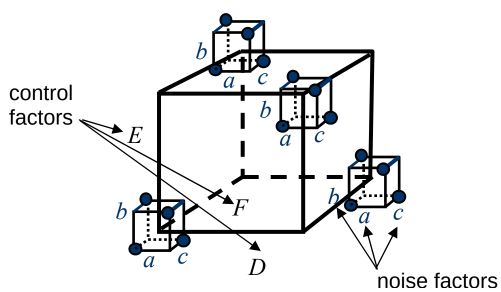
1) the experimental designs;
2) the universal use of signal to noise ratios as the choice of criteria;
3) the "one shot" format of the method that does not leverage the sequential nature of investigation [18].

Developments along these three lines have proceeded in parallel.

One important aspect of Taguchi's method was use of orthogonal arrays that provide advantages of balanced search and efficiency in the presence of experimental error. In most applications of Taguchi methods, a crossed array arrangement is suggested in which control factors are allocated to an "inner" array and noise factors are allocated to an "outer" array. The "outer array" describes the way that the noise factors varied systematically (they are brought under the experimenter's control for the purpose of the experiment). In a "crossed" array, the same pattern of noise factor variations are induced for each combination of control factor settings in the inner array. Each row of the noise array enables the experimenter to compute a signal to noise ratio. The effects of the control factors on the signal to noise ratio are analyzed and used in an effort to maximize robustness of the system. Taguchi methods emphasize the definition of an "ideal function" of the system and the use of dynamic signal to noise ratio wherever possible in order to measure robustness across a range of signal inputs to the system. To reduce the experimental effort, Taguchi methods may include “compound noise” in which the outer array is collapsed into just two settings in which all the noise factors are changed simultaneously.

A depiction of a very simple crossed array is provided in Figure 16. In the depiction, an L4 or 2_{III}^{3-1} array is used for both the inner and outer arrays. In practice, the inner and outer arrays are usually not the same, but using the smallest available alternative for both inner and outer arrays facilitated this graphical depiction. In particular, the inner array is usually a larger, more complex arrangement such as an L18 which allows exploration of more control factors and more levels of the control factors.

Figure 16 A cross array for robust design



One of the main issues raised has been the use of crossed arrays as compared to combined arrays (in which control and noise factors are varied jointly according to a single plan). A key advantage of the combined array approach is that it provides flexibility to for the designer to rule out certain effects a priori and thereby accomplish savings in run size.

Welch et. al. [19] proposed that cross arrays be replaced by a combined array with both noise and control factors and applied this approach to computer experiments. This combined array approach was expanded upon and by Shoemaker et. al. [20] who applied the approach to physical experiments. Borror and Montgomery [21] proposed a combined array approach based on mixed resolution designs. Wu and Zhu [47] articulated the minimum J-aberration criterion for the selection of combined arrays based on principles of effect ordering in which different categories of effects are ranked a priori in terms of their likely importance to RPD. A key advantage of the combined array approach is that it provides flexibility to for the designer to rule out certain effects a priori and thereby accomplish savings in run size. In some cases, this might be accomplished on the basis of physical reasoning about the specific system being designed, but the minimum J-aberration criterion offers a more general-purpose approach. Many experts argue that the combined array is generally superior to the crossed array, for example Lorenzen expressed the view that crossed arrays could, in some sense, be characterized as "half as good" as combined arrays [18] and Wu and Hamada [6] have stated that some combined arrays are uniformly better than cross arrays of the same run size in terms of number of clear main effects and two-factor interaction. Some counterpoints have been made to the adoption of combined arrays. Kacker argued that combined arrays are overly sensitive to missing data and also may lead to difficulties when sources of variation are not included among the explanatory variables [13]. Wu notes that, if the cost of a noise run is much smaller than the cost of a control run, the cross array is quite economical [6]. Steinberg and Bursztyn [48] suggested that good designs must enable estimation of all control by noise interactions and therefore cross arrays constructed from Plackett-Burman designs offer a good approach with minimal run sizes.

A separate, but related issue, concerns analysis of the data. If a cross array is employed, some practitioners use summary statistics such as signal-to-noise ratios as a measure of robustness and then analyze the effects of control factors on these metrics. It has been demonstrated that signal-to-noise ratios can hide information about control by noise interactions [16, 20]. On the basis of such evidence, research proceeded on alternative analysis of data including "response modeling" [19], dual response approaches [49] and rejection of pre-decided criteria in favor of graphical analysis and discovery [16]. The design of the experiment and analysis of the data are interrelated issues. Employing combined arrays essentially rules out use of summary statistics such as signal-to-noise ratios and so "response modeling" is generally carried out meaning that the effects of noise factors and control by noise interactions are estimated and used to compute performance measures which are then used in selection of control factor levels.

Another concern is the degree to which adaptation can be leveraged within RPD. Although RPD methods are sometimes applied iteratively, the vast majority of published case studies employ a "one shot" approach in which the experiment is conducted as initially planned with only a confirmation experiment being dependent on what the data reveal. Box [36] argued that the iterative nature of scientific investigation is a key to successful outcomes and that tendency toward "one shot" procedures can result in less improvement of systems than would have been achieved by iterative procedures. An adaptive experimental procedure for seeking maxima was proposed by Friedman and Savage [7] which entailed repeatedly seeking maxima using single variables. Response Surface Methodology [50] can be used in RPD and includes substantial means for adaptation of experimental plans based on what data reveal. Pronzato [51] proposed an adaptive optimization procedure which corresponds to maximizing the sum of the current estimated objective and a penalization for poor estimation. Adaptive sequential procedures have been proposed specifically tailored to RPD [52]. Frey et. al. [11] showed that an adaptive one factor at a time (OFAT) method can be effective under some conditions. Frey and Wang [30] proved that adaptive OFAT exploits main effects with high probability and also exploits two factor interactions, especially if they are large. A modification of adaptive OFAT for application to RPD will be discussed in a later section.

The three preceding paragraphs describe proposed improvements to RPD methodology emerging in recent decades. This paper posits that there is substantial room for improvement in evaluation of these new RPD methods. It is possible that some developments based on theory will prove to be ineffective in practice. As illustration, consider the study by Kunert et. al. [22] who compared the effectiveness of combined array and cross array approaches. Both RPD methods were applied to a single engineering system, a sheet metal spinning process. A key result of the study was that the cross array found an effect on the variance that cannot be seen from the combined array. Extrapolating from their data, one may conclude that the cross array would have been more effective for bringing about the desired outcome (reduced dimensional variance of the sheet metal articles). Regarding data analysis, the results were less clear cut with both the "classical analysis" (using log of the empirical variance over the outer array as the response) and the "interaction analysis" (response modeling approach) providing some comparative advantages. This led the authors to suggest that the cross array should be used and that both analyses of its data should be carried out. A further conjecture made by Kunert et. al. is that the combined array approach failed in the study because it relies too much on effect sparsity.

In light of the evidence currently in the literature, there is no clear resolution regarding the choice of cross and combined arrays. The theoretical arguments in favor of combined arrays are persuasive, but the theory may be incomplete in important ways. The empirical study by Kunert et. al. [22] provides compelling support for cross arrays, but cannot be taken as conclusive. The paired comparison was carried out on a single engineering system and one cannot rule out the possibility that the observed results were unusual. One could, in principle, run similar paired comparisons on a much larger sample of engineering systems. However, this would be prohibitively expensive. A more economical evaluation technique seems to be needed to bridge the gap between theory and large empirical studies. This is the principal motivation for this study.

4.4 Objectives and Methods

This section describes an approach employing computational simulation to compare the effectiveness of robust design methods. The approach described here relies on statistical characterization of the robust design methods by probabilistic simulation. Therefore, each robust design method must be repeated on a large enough number of individual systems to determine its average performance and the variations about the average. A key to the approach is appropriate modeling of interactions among the system input variables. The approach has four main steps:

1) Instantiate models of multiple engineering systems that will be subject to robust parameter design.
2) For each system, simulate every robust design method to determine the control factor settings preferred according to that method.

3) For each system/method pair, perform a confirmation experiment to determine the actual variance of the response at the chosen control factor settings.
4) For each method, analyze the data across all instantiated systems to determine the mean improvement and inter-quartile range for the improvement.

Each of these steps requires substantial elaboration as presented in the following subsections.

4.4.1 The Model and the System Regularities

In order to carry out the approach described herein, it is necessary to have a large number of engineering systems which are subject to robust design experiments. One approach would be to create a database of thousands of simulations of actual engineering systems. However, most simulations used to design large-scale engineering systems require hours, days, or more to produce realistic estimates of system performance at a single design point. The method described here requires thousands of such performance estimates for each system. A much faster approach is needed for the present purpose.

Another approach is to assemble a collection of data from experiments on a large number of authentic engineering systems. However, this approach did not allow enough flexibility for the present purposes or enable a study of systems with adequate scale. Systems with seven or more factors are rarely documented in adequate detail to enable the kind of approach presented here.

Motivated by the considerations described previously, we developed a way to create multiple instances of computationally efficient simulated systems which are realistic in certain critical regards. In particular, the interactions between control factors and noise factors must be modeled realistically because these provide opportunities for improved robustness. Other system interactions such as those among noise factors are also important because they create difficulties that may adversely affect the robust design methods being evaluated. Thus, for the present purpose, the structure of all interactions must be modeled in an appropriate way. As discussed in Chapter 3, actual engineering systems are verified and quantified to exhibit sparsity of effects, hierarchy, and inheritance. These properties are therefore central to the model proposed here. The proposed hierarchical probability model is expressed in Section 2.4.3, Equation 10 through 19. The model is reproduced below.

$$y(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j>i}}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j>i}}^{n} \sum_{\substack{k=1 \\ k>j}}^{n} \beta_{ijk} x_i x_j x_k + \varepsilon \tag{10a}$$

$$x_i \sim NID(0, \sigma_N^2) \;\; i \in 1 \ldots m \tag{11a}$$

$$x_i \in \{+1, -1\} \;\; i \in m+1 \ldots n \tag{12a}$$

$$\varepsilon \sim NID(0, \sigma_\varepsilon^2) \tag{13a}$$

$$f(\beta_i | \delta_i) = \begin{cases} N(0,1) & \text{if} \quad \delta_i = 0 \\ N(0,c) & \text{if} \quad \delta_i = 1 \end{cases} \tag{14a}$$

$$f(\beta_{ij} | \delta_{ij}) = \begin{cases} N(0,s_1) & \text{if} \quad \delta_{ij} = 0 \\ N(0,c \cdot s_1) & \text{if} \quad \delta_{ij} = 1 \end{cases} \tag{15a}$$

$$f(\beta_{ijk} | \delta_{ijk}) = \begin{cases} N(0,s_2) & \text{if} \quad \delta_{ijk} = 0 \\ N(0,c \cdot s_2) & \text{if} \quad \delta_{ijk} = 1 \end{cases} \tag{16a}$$

$$\Pr(\delta_i = 1) = p \tag{17a}$$

$$\Pr(\delta_{ij} = 1 | \delta_i, \delta_j) = \begin{cases} p_{00} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} = 0 \\ p_{01} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} = 1 \\ p_{11} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} = 2 \end{cases} \tag{18a}$$

$$\Pr(\delta_{ijk} = 1 | \delta_i, \delta_j, \delta_k) = \begin{cases} p_{000} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} + \delta_\mathrm{k} = 0 \\ p_{001} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} + \delta_\mathrm{k} = 1 \\ p_{011} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} + \delta_\mathrm{k} = 2 \\ p_{111} & \text{if} \quad \delta_\mathrm{i} + \delta_\mathrm{j} + \delta_\mathrm{k} = 3 \end{cases} \tag{19a}$$

This hierarchical probability model allows any desired number of response surfaces to be created such that the population of response surfaces has the properties of sparsity of effects, hierarchy, and inheritance and has these properties to a degree under explicit control of the investigator. Equation 10a represents a response y whose standard deviation might be reduced via RPD methods. The independent variables x_i are both control factors and noise variables; no distinction is made in this notation except via the indices. Equation 11a shows that the first set of independent variables (x_1, x_2,… x_m) represent noise factors and are assumed to be normally distributed. Equation 12a shows that the other independent variables (x_{m+1}, x_{m+2},… x_n) represent control factors and are assumed to be two-level factors. The random variable ε represents the pure experimental error in the observation of the response which was assumed to be normally distributed.

The response surface is assumed to be a third order polynomial in the independent variables x_i. The coefficients β_i are the main effects. The coefficients β_{ij} model two-way interactions including control by control, control by noise, and noise by noise interactions. Similarly, the coefficients β_{ijk} model three-way interactions including control by control by noise and control by noise by noise interactions.

The values of the polynomial coefficients (β) are determined by a random process that models the properties of effect sparsity, hierarchy, and inheritance. Equation 5 determines the probability density function for the first order coefficients. Factors can be either "active" or "inactive" depending on the value (0 or 1 respectively) of their corresponding parameters δ_i. The parameter c determines the relative magnitudes of active and inactive effects. Similarly, Equations 15a and 16a determine the probability density function for the second order and third order coefficients respectively. In Equations 12a and 13a, the hierarchy principle is reflected in the fact that second order effects are only $1/s_1$ times as strong (on average) as first order effects and third order effects are only $1/s_2$ times as strong as second order effects.

Equation 17a is related to the sparsity of effects principle. There is a probability p of any main effect being active and smaller values correspond to sparser effects. Equations 18a and 19a enforce inheritance as well as effect sparsity. The likelihood of any second order effect being active is low if no participating factor has an active main effect and is highest if all participating factors have active main effects. Thus, generally one assigns values to the parameters of a hierarchical probability model so that $p_{11} > p_{01} > p_{00,}$ and so on.

The proposed model has several real-valued parameters which may have a significant effect on the inferences drawn from its use. One set of model parameters (denoted as the "fitted model" in Table 15) represents the authors' effort to fit of the hierarchical probability model to typical circumstances experienced in RPD. As described by Li and Frey [53] and in Chapter 3, a group of 46 full factorial experiments were collected from technical journals (including 2^3, 2^4, 2^5, and 2^6 designs). In this data set, 41% of main effects were active according to the step-down

Lenth method using the SME criterion [39]. Therefore, the parameter p was set to 0.41 in the fitted model. The other probabilities in Equations 18a and 19a were similarly estimated based on the proportion of interaction effects of each type that were deemed as active in the set of full factorial experiments.

The parameter c was set to ten following the example in Chipman et. al. [37]. The parameter σ_N was set to 0.5 so that the range of noise factors in the experiment is somewhat smaller than the range of control factors. The parameter σ_ε was set to be substantially smaller than a typical main effect based on the observation that in electrical and mechanical sciences where RPD is most often applied, measurement abilities often far exceed those available within the chemical and life sciences [14] and gage R&R studies will sometimes ensure that pure error is less than 1/10 of the effects to be measured. The parameters s_1 was set so that two factor interactions would be smaller than main effects as indicated by our data and s_2 was similarly set so that three factor interactions would be smaller than two-factor interactions.

The subsequent rows in Table 15 are variants upon the fitted model intended to explore the influence of the model parameters on our results. The model denoted as "fitted, high ε" raises the degree of pure experimental error leaving all other parameters the same as in the fitted model. The model denoted as "sparse" makes all the effects less likely to be active while retaining hierarchy and inheritance. The model denoted as "strong hierarchy" makes all three-factor interactions inactive and also makes inactive three factor interactions particularly small.

Table 15 Model variants and associated sets of parameters

	p	p_{11}	p_{01}	p_{00}	p_{111}	p_{011}	p_{001}	p_{000}	c	σ_N	σ_ε	s_1	s_2
Fitted model	0.41	0.33	0.04	0.005	0.15	0.07	0.03	0.01	10	0.5	0.1	3.6	7.3
Fitted model, high ε	0.41	0.33	0.04	0.005	0.15	0.07	0.03	0.01	10	0.5	1	3.6	7.3
Sparse	0.2	0.1	0.01	0.005	0.1	0.01	0.01	0.01	10	0.5	0.1	3.6	7.3
Strong Hierarchy	0.41	0.33	0.04	0.005	0.0	0.0	0.0	0.0	10	0.5	0.1	3.6	21

The hierarchical probability models described have many limitations as means for evaluating RPD methods. Real engineering systems have responses that are often poorly fit by polynomials. In addition, it should be noted that the model as defined (10a-19a) restricts RPD as a search among discrete levels, although many RPD applications may include interpolation and extrapolation of continuous factors. Due to these limitations, conclusions based on this model should be assessed by checking them against other forms of evidence such as case studies based on engineering simulations and from field research with physical experiments.

4.4.2 Instantiate Multiple Response Surfaces

The proposed evaluation process is meant to indicate typical results and ranges of results across many uses of the method. Therefore, each robust design method to be evaluated must be repeated on a large number of different responses. In the proposed approach, these subjects are response surfaces created using the hierarchical probability model described in Section 4.4.1. The objectives of the study will determine the number of control factors and noise factors to be considered. The sample size in the study can then be determined based on an estimate of the variance of the primary variable and the confidence level to be used. Then the parameters δ may be formed for each system response using pseudo-random number generators. Based on the values of δ, the values of the polynomial coefficients β may be formed using pseudo-random number generators.

4.4.3 Simulate the RPD Methods

Every RPD method to be evaluated must be coded as an algorithm executable by a computer. Every method to be evaluated must also include a description of what simulated experiments are to be performed (i.e., at what control and noise factor values is the response y to be sampled). Every method to be evaluated must also include a description how data are analyzed and how the "optimized" control factor settings are determined.

Once the RPD methods are executable by computer, each method applied to each response surface. This approach is known as a paired comparison design in statistics and known as a crossover study in clinical medical practice. This design increases the power of the study substantially and therefore reduces the required sample size. In clinical practice, the crossover design introduces possibilities for bias due to carryover effects. However, in this computer simulation, carryover effects can easily be eliminated by resetting all the appropriate variables in the analysis.

4.4.4 Evaluate the Primary Variable

In this step, a variable is estimated which serves as a primary measure of the effectiveness of each method. In RPD, any particular method will result in a set of control factor settings which are predicted to be optimal. In our notation these, predicted optimal control factor settings are a vector of values at coded levels $x_i \in \{+1,-1\}$ where $i \in m+1 \ldots n$. Most methods provide a prediction of a robustness measure at the optimized condition. The predicted value and observed value frequently differ; therefore many methods call for a confirmation experiment in which the performance at the selected control factor settings is evaluated experimentally at multiple noise factor values. In the process proposed here, the confirmation experiment can be simulated by computing the exact solution for the transmitted variance of the response σ^2. This is possible in this model-based approach because the true values of parameters β are available. The transmitted variance is due to the noise factors which are a subset of the x_i. Given the response is a polynomial (Equation 10a), and given the assumption that the noise factors are independent, normally distributed, and have variance σ_N^2, it can be shown that

$$\sigma^2(x_{m+1}, x_{m+2}, \ldots, x_n) =$$

$$\sigma_N^2 \sum_{i=1}^{m} \left[\beta_i + \sum_{\substack{j=m+1 \\ j>i}}^{n} \beta_{ij} \cdot x_j + \sum_{\substack{j=m+1 \\ j>i}}^{n} \sum_{\substack{k=m+1 \\ k>j}}^{n} \beta_{ijk} \cdot x_j \cdot x_k \right]^2 + \sigma_N^4 \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j>i}}^{m} \left[\beta_{ij} + \sum_{\substack{k=m+1 \\ k>j}}^{n} \beta_{ijk} \cdot x_k \right]^2 + \sigma_N^6 \sum_{i=1}^{m} \sum_{\substack{j=1 \\ j>i}}^{m} \sum_{\substack{k=1 \\ k>j}}^{m} \beta_{ijk}{}^2 \tag{57}$$

Equation 57 is an extension of the standard variance model used, for example, in Morrison [54] or Borror and Montgomery [21]. This extension includes the influence of three-factor interactions which was critical since the influence of these effects appears to be of substantial practical importance in RPD as revealed by Kunert et. al. [22].

4.4.5 Analyze and Present the Data

The data generated by the simulations must be analyzed to draw inferences or to support decision making. Because the sample sizes can easily be increased in the computer simulation, most any small difference between two methods can be made statistically significant by raising the number of replications. This does not however guarantee that the result is reliable or of practical significance. The reliability of the results from this technique are strongly affected by the form of the model and the parameters used. Therefore it is advised that different variants of the model be run to determine the sensitivity of the results to the assumptions of the study.

The amount of performance improvement delivered by an RPD method can be communicated through a measure of central tendency of the primary variable (e.g., transmitted variance). However, a measure of central tendency is probably not adequate. Engineers must be concerned with the risk that a poor outcome may arise from a method, even if the method is usually effective. The designer may prefer a method that is more consistent in its results even if it is worse on average. Therefore measures of the range of results should also be presented.

An engineering designer often seeks to balance the total cost of carrying out RPD method against the benefits derived, therefore indications of cost should be displayed. Run size is one important measure that is commonly used, but it is not usually sufficient. Wu advocated that the number of "control runs" and the number of "noise runs" should be considered [18]. In some cases, the cost of changing control factor levels is an important cost driver, such as when different parts must be swapped in and out of an engineering system to bring about the level changes. In such scenarios, replacing one part may be much less time consuming and expensive than replacing several parts.

4.5 Case Study I – Adaptive OFAT in Robust Design

4.5.1 Selected Robust Design Methods

The purpose of this section is to evaluate several RPD methods of practical interest. In this study, all of the response surfaces had six control factors and three noise factors. We chose this scenario so that direct comparison could be made to the study by Kunert et. al. [22]. The RPD methods evaluated in the study are listed below.

$2_{III}^{6-3} \times 2_{III}^{3-1}$ with response modeling -- This method was intended to be similar to that employed by Kunert et. al. [22] in carrying out a cross array design and analyzing the results by response modeling. A fractional factorial 2_{III}^{6-3} inner array of control factors was crossed with a fractional factorial 2_{III}^{3-1} outer array of noise factors to give a 32-run design. The data from the observations were used to estimate all noise main effects and control by noise interactions. Based on these estimated effects, the transmitted variance was computed using Equation 57 for every possible combination of the control factor settings. The combination of control factor settings providing the lowest predicted transmitted variance was selected. If this method is carried out with a restriction on randomization of run order (so that the noise array at each control factor setting is contiguous), then 21 changes in control factor levels are required on average (seven changeovers between runs in the inner array, each requiring three level changes on average between subsequent runs).

$2_{VI}^{6-1} \times 2_{III}^{3-1}$ with response modeling -- This process was intended to provide an indication of how much value is derived by increased resolution of the inner array. A resolution *VI* fractional factorial 2_{VI}^{6-1} array of control factors was crossed with a fractional factorial 2_{III}^{3-1} outer array of noise factors. The data from the design were used to estimate all noise main effects and control by noise interactions. Based on these estimated effects, the transmitted variance was computed using Equation 57 for every possible combination of the control factor settings. The combination of control factor settings providing the lowest predicted transmitted variance was selected.

$2_{III}^{6-3} \times 2_{III}^{3-1}$ with classical analysis -- This process was intended to be similar to that employed by Kunert et. al. in carrying out a crossed array design and analyzing the results by what they refer to as "classical analysis." A fractional factorial 2_{III}^{6-3} inner array of control factors was crossed with a fractional factorial 2_{III}^{3-1} outer array of noise factors to give the same 32-run design as in Kunert et. al. [22]. The data from the design were used to calculate the control factor effects on the log of the transmitted variance. This measure sometimes called a type II signal-to-noise ratio [25]. The control factor settings were selected based on the sign of the factor's effects on the log of the transmitted variance.

$2_{III}^{6-3} \times 2_{III}^{3-1}$ with both analyses -- This process was intended to correspond to the suggestion by Kunert et. al.[22] that both response analysis and classical analysis should be carried out on the data from a cross array. In simulating this process, both analyses described above were conducted and each led to a set of predicted control factor settings. These two settings typically were different in at least one factor, so two different confirmation experiments were required to make a final determination of the optimized settings. Whichever control factor setting provided the lower confirmed variance was selected. Thus, the overall experiment requires 32 runs for the cross array and typically eight runs for the two confirmation experiments or 40 runs in all.

2^{9-4} with largest noise clear-- A 32 run combined array approach was employed as described by Kunert et. al. assigning the noise factor with the largest main effect to the array so that its interactions with control factors were clear. The combined array was executed and the resulting data were used to calculate the main effects of the noise factors and control by noise interactions. Based on these estimated effects, the control factor settings were selected.

2^{9-4} with minimum *J*-aberration -- A 32 run combined array approach was selected on the basis of minimum *J*-aberration as described by Wu and Zhu [47]. The combined array was executed and the resulting data were used to calculate the main effects of the noise factors and control by noise interactions. Based on these estimated effects, the control factor settings were selected.

$aOFAT \times 2_{III}^{3-1}$ p_{guess}= 50% -- Adaptive one-factor-at-a-time experimentation was applied to the control factors as described in Frey and Wang [30]. An extension of this approach was used here as described in Figure 17. Adaptive OFAT was crossed with a resolution *III* outer array of noise factors so that, for each setting of the control factors, a fractional factorial 2_{III}^{3-1} outer array of noise factors was run and the transmitted variance of the observations was taken as a response to be reduced. A starting point set of control factor levels was selected at random. Any control factor changes reducing the observed variance were retained and otherwise the changes were revered before proceeding. The overall process requires 24 runs and about nine changes in control factor levels since each factor is toggled once and the level changes are reversed about half of the time.

$aOFAT \times 2^{3-1}_{III}$ p_{guess}= 75% -- Adaptive one-factor-at-a-time experimentation was applied as depicted in Figure 17 and is identical to the process described above except that the starting point was selected so that each factor was set independently to have a 75% chance of matching the setting with the lowest transmitted variance among all 64 combinations of factor levels. This is intended to represent a situation in which the experimenter is uncertain, but not entirely indifferent, about the best starting point for the adaptive process. This models a situation wherein the experimenter has some *a priori* basis for judging one level of each factor as more promising than the alternative. The overall process given p_{guess}= 75% requires essentially the same costs as the method with p_{guess}= 50% except that about 10.5 changes in control factor levels will be needed on average since all six levels are changed initially and these level changes are reversed about 75% of the time.

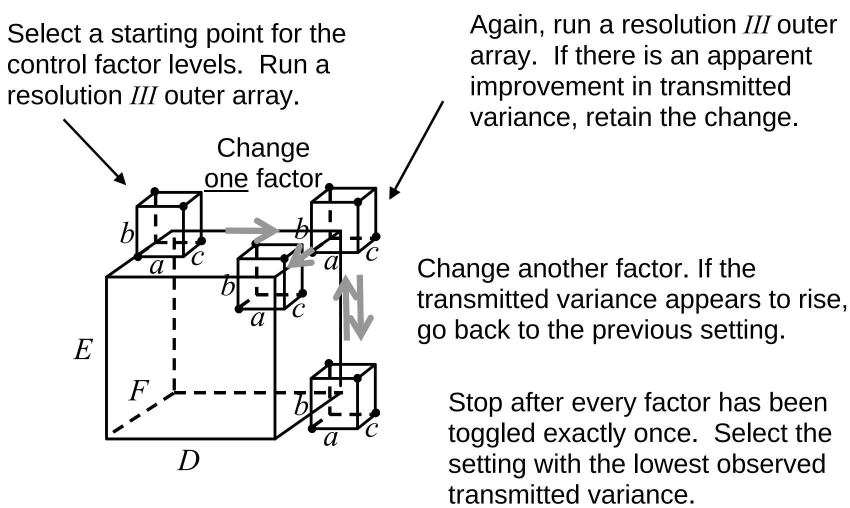


Figure 17 The adaptive one factor at a time method crossed with a resolution III outer array. This method is denoted as $aOFAT \times 2^{3-1}_{III}$ in this case study. Noise factors *a, b,* and *c* are varied according to a factorial design and control factors *D, E,* and *F* are explored sequentially to seek lower variance in the observed response.

Cost indicators for each of these methods are summarized in Table 16. Run size is a common indicator of resources required since each run requires time to carrying out an experimental protocol and make observations. The number of control factor changes to be made is another critical indicator since changes in factor levels between subsequent experiments frequently require labor to install components of different types. The number of noise factor changes can also be important, but is often less costly than control factor changes. The methods are arranged so that, accounting for all the cost indicators in a balanced fashion, the overall costs tend to decrease from left to right.

For each of these methods, a set of control factor settings emerge. To assess the outcome, the transmitted variance at the selected control factor settings was computed based on the actual model parameters β rather than the estimates which would have been affected by the experimental design and pure experimental error. This is in some ways similar to a confirmation experiment, but it is not subject to the potential errors in confirmation experiments. The percentage reduction from the average transmitted variance across the entire design space for that system was computed and recorded.

Each method was applied to 10,000 instances of response surfaces sampled from each model variant. For each method/model variant pair, the average percent reduction in transmitted variance across the 10,000 responses was computed as reported in Table 17. The inter-quartile ranges of percent reduction are reported in Table 4 giving an indication of the repeatability of the methods across different engineering systems to which they may be applied

Table 16 Cost indicators for the various methods considered in this case study.

	$2_{VI}^{6-1} \times 2_{III}^{3-1}$	2^{9-4}		$2_{III}^{6-3} \times 2_{III}^{3-1}$			$aOFAT \times 2_{III}^{3-1}$	
	response modeling	largest noise clear	minimum J-aberration	both analyses	Classical analysis	Response modeling	p_{guess} = 75%	p_{guess} = 50%
Number of experimental runs	128	32	32	40	32	32	28	28
Number of control factor level changes	384	96	96	27	21	21	10.5	9
Number of noise factor level changes	255	63	63	71	63	63	55	55

Table 17 Mean percent reduction in transmitted variance achieved by various methods applied to various models.

	$2_{VI}^{6-1} \times 2_{III}^{3-1}$	2^{9-4}		$2_{III}^{6-3} \times 2_{III}^{3-1}$			$aOFAT \times 2_{III}^{3-1}$	
	response modeling	largest noise clear	minimum J-aberration	both analyses	Response modeling	Classical analysis	p_{guess} = 75%	p_{guess} = 50%
Fitted model	65%	20%	24%	51%	30%	39%	60%	55%
Fitted model, high ε	55%	17%	22%	46%	21%	30%	47%	40%
Sparse	57%	17%	23%	44%	31%	34%	52%	47%
Strong Hierarchy	60%	32%	38%	57%	56%	50%	57%	54%

Table 18 Inter-quartile range of percent reduction in transmitted variance achieved by various methods applied to various models.

	$2_{VI}^{6-1} \times 2_{III}^{3-1}$	2^{9-4}		$2_{III}^{6-3} \times 2_{III}^{3-1}$			$aOFAT \times 2_{III}^{3-1}$	
	response modeling	largest noise clear	minimum J-aberration	both analyses	Respon-se model-ing	Classi-cal analy-sis	$p_{guess} = 75\%$	$p_{guess} = 50\%$
Fitted model	46 to 87%	-1 to 53%	3 to 59%	28 to 77%	9 to 65%	14 to 69%	39 to 82%	34 to 77%
Fitted model, high ε	33 to 80%	-2 to 52%	0 to 54%	24 to 72%	4 to 57%	8 to 63%	27 to 73%	17 to 68%
Sparse	32 to 83%	1 to 45%	4 to 48%	20 to 70%	9 to 53%	12 to 58%	28 to 76%	24 to 69%
Strong Hierarchy	31 to 83%	13 to 61%	16 to 66%	31 to 84%	31 to 83%	25 to 74%	31 to 80%	28 to 76%

4.5.2 Discussion of the Case Study I Results

The results of this study are summarized in Tables 16, 17 and 18 wherein the rows denote the model variants and the columns denote the methodological alternatives assessed. In Table 17, the value within each cell is the mean percentage by which the transmitted variance was reduced across the population of 10,000 response surfaces sampled from that model and then improved by that method. In Table 18, the two values within each cell indicate the inter-quartiles of percentages by which the transmitted variance was reduced across the population.

This study generally supports the contention that minimum *J*-aberration is a good criterion for selection of combined arrays as proposed by Wu and Zhu [47]. This result was consistent, although not large, across all the model variants we considered. This is interesting since the alternative selection criterion relied upon accurate *a priori* selection of the largest noise main effect which would seem to be a significant advantage in planning for RPD.

Another key result is corroboration of Kunert et al. [22] regarding comparative advantages of crossed arrays versus combined arrays. Based on the "fitted model", cross arrays do appear to give better results than combined arrays regardless of how the data are analyzed and regardless of which criterion is used to select the combined array. This result was consistent across all the model variants considered here. A principal difficulty experienced with the combined arrays was that control by control interactions influenced the estimates of control by noise interactions. A product structure prevents such influences which, according this study, leads to better results. A similar issue concerning combined arrays was raised by Shoemaker et. al. [20] who noted the reliance of combined arrays on "how well the model fits" and warned that "the combined array experiment may lead to control-factor settings that actually increase variability". This was observed in this study as shown in Table 18. Some caution is required in that this study does not consider adjustment to the target assuming either that this is not needed or that it can be accomplished equally well by all the methods. Thus, combined arrays may have advantages not well evaluated by this study.

An advantage of the model-based assessment is that one may explore the influence of one's assumptions on selected conclusions. In particular, Kunert et. al. suggested that combined arrays failed to perform as well as crossed arrays because they rely too strongly on effect sparsity. In this study, making the effects more sparse did not improve the performance of combined arrays very much. On the other hand, a greater degree of hierarchy appeared to have a substantial influence on the effectiveness of combined arrays. The model denoted "strong hierarchy" included significant two-factor interactions but no large three factor interactions. Under these conditions, combined arrays performed better, but still not as well as cross arrays.

A surprising result of this study is the comparative performance of the two data analysis methods applied to crossed arrays. For the fitted model, a classical analysis which corresponds to use of Taguchi's type *II* signal to noise ratio provided better outcomes than response modeling. This is observed despite the fact that such analysis sometimes masks important control by noise interactions. The reason may be, as observed in the sheet metal spinning case study, that the classical analysis will sometimes exploit the benefit of control by control by noise interactions even though these interactions cannot be resolved by this approach. Reinforcing this hypothesis, the advantage of the classical analysis disappeared for the strong hierarchy model. This study therefore tends to reinforce the suggestion made by Kunert et. al. that there exists an "advantage of the cross array and classical analysis that should not be neglected: we manage to identify an important aspect of the process, without understanding what is really going on." This study supports the contention that classical analysis of the cross array provides benefits associated with effects which the cross array cannot resolve.

When using cross arrays, Kunert et. al. [22] recommended conducting both classical and response model analysis. This study confirms that this procedure provides substantially better outcomes than conducting either analysis alone. For all the models except "strong hierarchy", conducting both analyses provided 10% additional improvement (or more) as compared to the best single analysis. This benefit comes at the cost of conducting extra confirmation experiments which is a modest increment in resource demands.

In the view of the author, the most noteworthy results of this study concern the role of adaptation in RPD. For the fitted model, adaptive OFAT performed better than every alternative method that required similar resources including Kunert's suggested approach requiring an additional 12 runs. Across all the model variants, adaptive OFAT provided results comparable to the best alternative using similar resources and within a modest increment of the $2_{VI}^{6-1} \times 2_{III}^{3-1}$ approach which requires more than four times as many runs. Adding a larger degree of experimental error did not degrade the performance of adaptive OFAT substantially more than other methods. It was found that $2_{III}^{6-3} \times 2_{III}^{3-1}$ with both analyses provided a small advantage over adaptive OFAT with p_{guess}=50% for the fitted, high ε model and for the strong hierarchy model. Given the slightly more optimistic assumption that p_{guess}=75%, adaptive OFAT is superior to every alternative RPD method with similar resource demands and this result persists across every set of assumptions modeled. These results generally run counter to the widely held view that factorial design beats one factor at a time experimentation for RPD, as expressed for example by Kacker [13].

The relatively good performance of adaptive OFAT in this case study will appear counter-intuitive to many and therefore requires some explanation. Frey and Wang [30] show that adaptive OFAT applied without the cross array will exploit main effects and will also exploit two factor interactions more often than not and will exploit the largest interactions with high probability. If adaptive OFAT is crossed with a resolution *III* array of noise factors, it follows that control by noise interactions are exploited and that control by control by noise interactions will be exploited with high probability especially when they are large. The design does not enable resolution of these effects. Thus, some improvements in robustness are attained without knowing precisely from whence they came although one can ascertain which control factor was linked to the improvements. Our conclusion is that adaptive OFAT offers an advantage in robustness improvements attained, but with an attendant drawback regarding the information provided.

4.5.3 Suggestions for Future Research Following Case study I

The conclusion that adaptive OFAT crossed with resolution III arrays provides outcomes superior to other RPD methods should be put to additional tests. Specifically, we propose that the adaptive OFAT process described here should be field tested against the best alternatives available. Paired comparisons using physical experiments should be conducted following the approach in Kunert [22]. In addition, case studies with realistic computer models of engineering systems should be carried out as these will be more readily replicated by multiple independent investigators.

Table 18 provides an interesting opportunity for further investigation. The authors are unaware of any published reports of typical ranges of robustness improvement drawn from an unbiased set of industry applications of RPD. The ranges reported in Table 18 therefore represent testable empirical claims. For example, based on the crossed array with classical analysis as applied to the fitted model, we estimate that the inter-quartile range of improvements due to this RPD method is 9 to 65%. This implies that about 25% of all applications provide either no improvement or very little improvement. Testing this claim requires either an appropriately controlled field study with adequate sample size, or a record of past applications that does not censor the poorest outcomes.

The model-based evaluation of RPD presented here places emphasis exclusively on the outcomes of the experimentation method in terms of the performance of the engineering system. The evaluation method does not attempt to account explicitly for the value of scientific learning. Such considerations may have a significant impact on the long-term value provided by industrial experimentation. The case study presented here suggests that, by using adaptive OFAT, costs might be reduced and larger improvements attained (in the short term) by trading off explicit knowledge gained. Other methods may promote more learning and discovery than adaptive OFAT and therefore result in greater long-term benefits. These hypotheses should be tested using field data with human subjects or scientifically validated models of human learning. Such studies hold out the exciting prospect of uniting diverse, but interrelated disciplines such as DOE and cognitive psychology.

4.6 Case Study II – Compounding Noise Strategy in Robust Design

4.6.1 Select Parameters of the Hierarchical Probability Model

The hierarchical probability model described in Section 4.4.1 was used to create the simulated engineering systems for this case study. The systems in this case study all have seven noise factors and seven control factors, therefore m=7 and n=14. These values were chosen because they represent a reasonable number of factors that might be considered in industrial practice of robust design. This choice results in a potentially very complex model including 469 coefficients. Of these, 364 coefficients represent three-way interactions, but the vast majority of these will be inactive due to the assumption of hierarchy and inheritance.

The model has several real valued parameters which may have a significant effect on the inferences drawn from its use. To provide a balanced view, six different sets of parameter settings were used as shown in Tables 19 and 20.

Table 19 Sets of model parameters considered in the case study

	c	s1	s2	w1	w1
Basic WH	10	1	1	1	1
Basic low w	10	1	1	0.1	0.1
Basic 2nd order	10	1	0	1	1
Fitted WH	15	1/3	2/3	1	1
Fitted low w	15	1/3	2/3	0.1	0.1
Fitted 2nd order	15	1/3	0	1	1

Table 20 Additional model parameters for each set considered in the case study

	p	p11	p01	p00	p111	p011	p001	p000
Basic WH	0.25	0.25	0.1	0	0.25	0.1	0	0
Basic low w	0.25	0.25	0.1	0	0.25	0.1	0	0
Basic 2nd order	0.25	0.25	0.1	0	N/A	N/A	N/A	N/A
Fitted WH	0.43	0.31	0.04	0	0.17	0.08	0.02	0
Fitted low w	0.43	0.31	0.04	0	0.17	0.08	0.02	0
Fitted 2nd order	0.43	0.31	0.04	0	N/A	N/A	N/A	N/A

The Basic WH model is based on the prior parameters used in Bayesian model selection by Chipman, et. al [37]. Two variants were developed from that basic model. The low w variant accounts for the fact that, in robust design, the control factors are generally explored over a wider range than the noise factors. The 2^{nd} order variant zeros out the coefficients of all the three way interactions.

The fitted weak heredity model (fitted WH), was developed specifically for use in this case study. The parameters of the model were selected based on their fit to a set of empirical data. A group of 62 experimental data sets were collected which have 2^4 full factorial designs. In this data set, 43% of main effects were active according to the step-down Lenth method [39]. The other probabilities were estimated in a similar way.

The estimation of the other parameters (c, s_1, etc.) was based exclusively on the four factor experiments 2^4 of the 62 responses. Firstly, the data were normalized so that we could meaningfully compare values across systems. The responses from each experiment were transformed so that the minimum observation was zero and the maximum observation was 100. The factor effects were then computed from the normalized data. The value of s_1 was estimated by computing the ratio of the standard deviation of all main effects and the standard deviation of all two-way interactions. The value of s_2 was estimated the same way but by comparison with three-way interactions. Histograms of factor effects were formed for the collection of real systems and for a set of 1000 systems instantiated from the WH model. The shape of the distribution was used to adjust the parameter c. If the setting of c is too low, the histogram of factor effects has overly thick tails. This led to adjustment of c from 10 to 15 which provided a more reasonable fit of the model to the data (as shown in Figures 18 and 19).

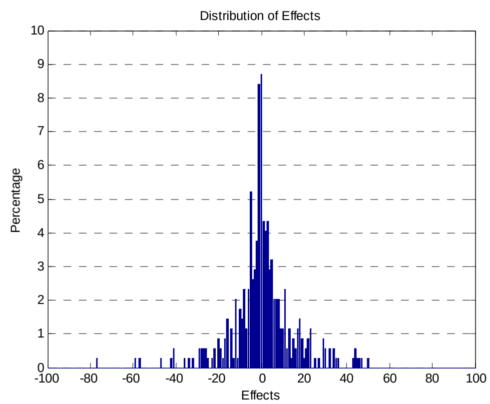


Figure 18 The distribution of factor effects from real systems

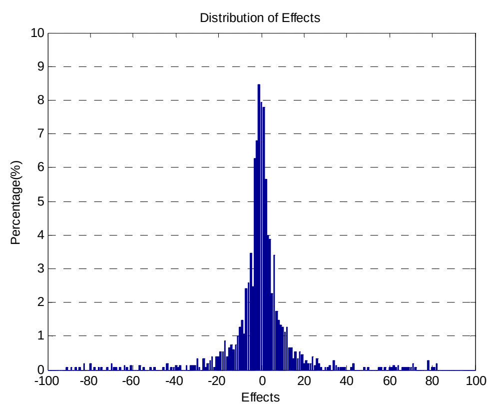


Figure 19 The distribution of factor effects from 1000 simulated systems sampled from the fitted weak heredity model with c=15

4.6.2 Robust Design Methods to Be Evaluated

Six different approaches to robust design were evaluated. They all involve a crossed array strategy of one sort or other, but are differentiated by the resolutions of the inner and outer arrays.

$2^7 \times 2^7$ -- A full factorial 2^7 inner array of control factors was crossed with a full factorial 2^7 outer array of noise. This approach is inordinately expensive, requiring 16,348 experiments, but provides resolution to estimate all 469 coefficients in the model. Based on these parameters, the standard deviation was estimated and the control factor settings were optimized based on these estimates. This design provides a baseline value since the standard deviation optimized by this approach will be the lowest possible value within the given discrete space of control factor settings.

$2^7 \times 2^{7-4}_{III}$ -- A full factorial 2^7 inner array of control factors was crossed with a fractional factorial 2^{7-4}_{III} outer array of noise. The data from the design were used to calculate all noise main effects, control by noise interactions, and control by control by noise interactions. Based on these parameters, the standard deviation was estimated based on Equation 12 and the control factor settings were optimized based on these estimates.

$2^7 \times CN_s$ -- A full factorial 2^7 inner array of control factors was crossed with a compound noise factor. First a 2^{7-4}_{III} design was performed on the noise factors alone to estimate their main effects. Then level of each noise factor within the compound noise factor was set as the sign of its main effect estimate. The crossed array was then executed. The data from the design were used to calculate the main effect of the compound noise and control by compound noise interactions. Based on these parameters, the standard deviation was estimated based on Equation 12 and the control factor settings were optimized based on these estimates.

$2_{III}^{7-4} \times 2_{III}^{7-4}$ -- A fractional factorial 2_{III}^{7-4} inner array of control factors was crossed with a fractional factorial 2_{III}^{7-4} outer array of noise. The data from the design were used to calculate all noise main effects and control by noise interactions. Based on these parameters, the standard deviation was estimated based on Equation 12 and the control factor settings were optimized based on these estimates.

$2_{III}^{7-4} \times CN_S$ -- A fractional factorial 2_{III}^{7-4} inner array of control factors was crossed with a compound noise factor. First a 2_{III}^{7-4} design was performed on the noise factors alone to estimate their main effects. Then level of each noise factor within the compound noise factor was set as the sign of its main effect estimate. The crossed array was then executed. The data from the design were used to calculate the main effect of the compound noise and control by compound noise interactions. Based on these parameters, the standard deviation was estimated based on Equation 12 and the control factor settings were optimized based on these estimates.

$2_{III}^{7-4} \times CN_R$ -- A fractional factorial 2_{III}^{7-4} inner array of control factors was crossed with a compound noise factor. The levels of the compound noise factor were set to +1 or -1 at random. The crossed array was then executed. The data from the design were used to calculate the main effect of the compound noise and control by compound noise interactions. Based on these parameters, the standard deviation was estimated based on Equation 12 and the control factor settings were optimized based on these estimates.

4.6.3 Results of Case Study II

The four step method described in Section 4.3 was applied to 36 pairings of six system types and six robust design methods. The results of the study are presented in Tables 21 and 22. The percentages depicted in the Table 21 are the expected value (averaged across the systems) of the percentage of confirmed improvement in the standard deviation of the response. The confirmed improvement is the confirmed standard deviation evaluated at the predicted optimum control factor settings. The percentage of confirmed improvement was defined so that 100% improvement implies the confirmed standard deviation is zero and 0% improvement implies that the confirmed standard deviation is no better than selecting the control factor settings at random from among the available discrete settings. The expectation is estimated by averaging the percentages across all 100 systems instantiated from that parameter set. The inter-quartiles are computed to provide an indication of the variability of the outcomes across different systems of the same type.

Table 21 Expected values of percent reduction in standard deviation for various robust design methods and system parameter sets

		Basic			Fitted		
Method	Experiments	WH	low w	2nd order	WH	low w	2nd order
$2^7 \times 2^7$	16,384	29%	54%	37%	31%	42%	27%
$2^7 \times 2_{III}^{7-4}$	1,024	16%	54%	33%	24%	41%	27%
$2^7 \times CN_s$	8+256	7%	14%	16%	18%	30%	21%
$2_{III}^{7-4} \times 2_{III}^{7-4}$	64	5%	6%	33%	9%	14%	27%
$2_{III}^{7-4} \times CN_S$	8+16	2%	6%	13%	13%	14%	21%
$2_{III}^{7-4} \times CN_R$	16	3%	3%	10%	4%	5%	11%

Table 22 Inter-quartile ranges of percent reduction in standard deviation for various robust design methods and system parameter sets

		Basic			Fitted		
Method	Experiments	WH	low w	2nd order	WH	low w	2nd order
$2^7 \times 2^7$	16,384	23 to 33%	46 to 62%	25 to 48%	23 to 40%	27 to 56%	14 to 36%
$2^7 \times 2_{III}^{7-4}$	1,024	8 to 23%	46 to 62%	22 to 45%	14 to 34%	27 to 55%	14 to 36%
$2^7 \times CN_s$	8+256	-4 to 18%	-1 to 28%	3 to 31%	7 to 28%	15 to 42%	7 to 31%
$2_{III}^{7-4} \times 2_{III}^{7-4}$	64	-4 to 15%	-1 to 26%	22 to 44%	0 to 22%	1 to 25%	14 to 36%
$2_{III}^{7-4} \times CN_S$	8+16	-7 to 12%	-9 to 20%	-15 to 24%	2 to 22%	0 to 25%	6 to 31%
$2_{III}^{7-4} \times CN_R$	16	-5 to 11%	-2 to 23%	-4 to 22%	-5 to 18%	-7 to 16%	-1 to 24%

4.6.4 Discussion of Case Study II

One salient feature of the data from this case study as presented in Table 21 and 22 is that the expected improvements in standard deviation are modest even in the best cases. The largest improvements in the standard deviation were for the basic weak heredity model with low w (the control factors varied over a wider range than the noises). The expected value of percent improvement was 54% with an inter-quartile range of 46-62%. In the worst cases, the mean improvements are very and the lower quartile is negative in some cases. This indicates that the confirmed standard deviation of the predicted "optimal" design was actually higher than the average standard deviation among all the available designs for that system.

The relatively low degrees of improvement require some discussion. The literature on robust design includes a many case studies in which large reductions in standard deviation were realized. There are several alternative explanations for the discord between the trends in the literature and the results in Tables 21 and 22. One possible explanation is that the parameters in the models are not realistic. It may be that, in robust design practice, engineers are usually able to improve the model additivity greatly by selection of variables and transformation of the response. Another possibility is that the limitation to two-level control factors hampers the effectiveness of robust design. Use of three-level control factors might improve the outcomes substantially. Another possible explanation is that the literature tends to document the most successful case studies and rarely documents unsuccessful attempts at robust design. It is possible that the literature represents substantially biased sample of the entire population of robust design applications.

It is significant that, for the basic and fitted 2nd order models, the $2_{III}^{7-4} \times 2_{III}^{7-4}$ method provides almost all the benefits of the full factorial design. This is a confirmation that in robust design it is best to focus on control by noise interactions as opposed to control by control or noise by noise interactions. However, the $2_{III}^{7-4} \times 2_{III}^{7-4}$ method does not fare as well in the third order models. The difference between the outcomes for the $2^7 \times 2^7$ and the $2_{III}^{7-4} \times 2_{III}^{7-4}$ derives primarily from the presence of control by control by noise interactions. The resolution III inner array cannot resolve these interactions and even confounds them with control by noise interactions.

Another interesting feature of these data is that on the fitted models, the $2_{III}^{7-4} \times CN_S$ method is very competitive as compared with the $2_{III}^{7-4} \times 2_{III}^{7-4}$. Further research is required to explain this phenomenon since the extreme conditions do not hold for any of the systems instantiated in this case study. It is possible that the results of further studies with this computational approach will show that compounding of noise does not require extreme conditions to be effective. It may be valuable to identify exactly what conditions should hold for the compound noise method to be effective.

4.7 Concluding Remarks of the Case Studies

We can frequently be surprised by what model-based assessments demonstrate. Although every conclusion emerging from the analysis is a necessary result of the assumptions in the model, the simulations reveal outcomes people cannot deduce unaided by computers. Therefore, model-based assessment of RPD methodology can fill a gap between conjectures based on theoretical understanding of statistics and practical field experience.

There are many specifics of the model, the evaluation technique, and the case study that can and should be subject to critique. The principal idea advanced in this paper is that model-based evaluations of RPD methods provide valuable insights. The approach enables assumptions about RPD such as effect hierarchy and sparsity to be quantified and expressed openly. Those who disagree with particular conclusions in this paper can identify the specific assumptions that may be at fault, implement changes in the assumptions, and observe the changes in the outcomes. It is proposed that such explorations can augment the scholarly investigation of RPD.

The case study in this paper suggests that incorporating a greater degree of iterative adaptation into RPD, even in a very simple way, has a large positive impact on the outcomes. The quantitative assessment presented here suggests that the benefits of adaptation are larger than those provided by many other methodological refinements in RPD that have been proposed in recent decades. We draw the conclusion that incorporating adaptation into RPD should be pursued more vigorously.

A final conclusion concerns the role of validation techniques in improving communication about RPD methods. One of the impressions made by the case study is that any of the RPD methods evaluated here can provide benefits rewarding those who invest time to learn and implement the methods. These benefits helped RPD to spread through many sectors of industry. On the other hand, Meyers and Vining noted in 1992 that only a small percent of American Companies were using statistical methods at all and expressed hope that there would be a profound increase is usage [18]. As applied to RPD, this hope has not been realized to a significant degree with industry application of RPD relatively flat over the past two decades.

The lack of profound increase may have been due to unrealistic expectations created when the methods were disseminated. Note that the case study suggests it is not uncommon for individual applications of popular RPD methods to provide no benefit at all. This observation seems to be consistent with the experiences of those applying RPD in industry. As observed by Kacker, the conditions needed for improvement by RPD "may not be rare, but certainly they are not universal" [13]. However, in published case studies on RPD, a reader will find many successful applications and hardly any failures. The publication process naturally favors significant, positive results [55]. In RPD, this may lead to a false impression. Based on these considerations, we propose that model-based evaluations including ranges of outcomes should be used in publication, education, and consulting. This practice would help set more realistic expectations and might prevent backlash against generally effective practices that occasionally yield disappointing results.

4.8 A Broader Discussion in Relation to Complex Engineering Systems

The evaluation technique and case study presented here may have some interesting implications for the emerging field of complex engineering systems. This broader discussion seems to fall naturally into three clusters which will be addressed in the following paragraphs:

1) The role of robust design in engineering systems;

2) Flexibility and the design process;

3) Possible directions for research methods;

Robust design is an important strategy for dealing with the uncertainties attendant in the design of complex engineered systems. If components and subsystems can be made robust to uncertain interface variables, then system integration is likely to proceed more smoothly. However, it is not at all obvious how best to deploy robust design in a Systems Engineering effort. Some robust design techniques provide good results reliably but at high cost. If these are used, then they can only be applied to the most critical components. Other robust design methods are simpler and less demanding of resources but occasionally result in disappointing outcomes. Such techniques can be taught to a large number of engineers and implemented (to some degree) on almost every component and subsystem. It falls upon the Systems Engineer to devise an overall strategy for deploying all available methods across the system. In order for Systems Engineers to perform that function, they need information about the pros and cons of every available technique. The literature on robust design includes arguments for and against different methods, but most of the arguments are made on the basis mathematical properties of the experimental designs. This paper presents an approach to evaluating robust design methods directly on the basis of their outcomes. I propose that information in this form is more valuable to the Systems Engineer than the kind of information previously available. Further, I believe that when Systems Engineers see that an adaptive OFAT approach will provide a 42% reduction in variance and that a much more expensive full factorial approach can only provide 58% (as shown in Table 21), there is likely to be a significant change in their behavior with more "quick and dirty" approaches deployed more broadly.

Flexibility is regarded as an important foundational issue in complex engineering systems. A central issue in designing any system is the tension between optimality for a fixed purpose and flexibility in the face of change. Most often, this tension is discussed vis-à-vis the flexibility/optimality of the product, but the tension is equally relevant to the design process. Should an enterprise plan a design process that will lead to the best outcomes assuming current knowledge of the design scenario, or employ a flexible design process that may not be ideally suited to any one condition but can adapt well in the face of new information that comes to light? This paper has shown that, in robust design, the value of adaptability and flexibility is so significant that, when accounted for, it overturns the conventional wisdom. No modern text on the subject of design or experiments or robust design suggests any valid role for an OFAT approach. Nevertheless, the data from these new simulations strongly suggests OFAT is a preferred design for an inner array. It appears that flexibility of the design process is far more valuable than previously acknowledged by the technical literature on design of experiments or robust design. Similar misconceptions may exist in other areas of Systems Engineering.

If the results of this paper are borne out, how can it be that the benefits of an OFAT approach to robust design have gone unnoticed given two decades of strong research efforts in this area? My view is that there has been an over-dependence on closed form analysis. Closed form proofs are a genuine and widely recognized sign of scholarly accomplishment. It is very difficult to write a proof showing that a complex adaptive process will have desirable properties. Therefore, adaptive processes for robust design have garnered little attention from academic communities. On the other hand, it is often easy to demonstrate the behavior of a complex process by computational simulation once a reasonable set of assumptions can be articulated and defended. One key set of assumptions for this study concerned the structure of interactions among variables in engineering systems. The development of the relaxed weak heredity model by Wu and Hamada was a critical enabler for the research presented here. It enabled the creation of multiple instances of reasonably realistic simulated engineering systems upon which robust design methods could operate. It may be that other aspects of the Systems Engineering process could also be simulated if the other reasonable assumptions can be identified and codified. If so, design process simulations like the ones discussed here could represent a promising research methodology for a wide range of Systems Engineering topics.

4.9 Summary

Following the system regularities investigated in Chapter 3, this chapter applies those regularities into system models and then uses the models to validate and compare different robust design methods. A new strategy is proposed for evaluating and comparing the effectiveness of robust parameter design methods. The hierarchical probability model is used to capture assumptions about robust design scenarios including effect sparsity, hierarchy, inheritance, and degree of experimental error. A process is presented employing this model to evaluate robust design methods. This process is then used to explore four topics of debate in robust design:

1) The relative effectiveness of crossed versus combined arrays;
2) The comparative advantages of signal-to-noise ratios versus response modeling for analysis of crossed arrays;
3) The use of adaptive versus "one shot" methods for robust design;
4) The expectation of improvements after applying robust designs.

In the first case study, it is shown that crossed arrays are preferred to combined arrays regardless of the criterion used in selection of the combined array. It is further shown that when analyzing the data from crossed arrays, signal-to-noise ratios generally provide superior results, but that response modeling should be used when the designer is highly confident that three-factor interactions are absent.

Most significantly, it is shown that using adaptive OFAT crossed with an orthogonal outer array results in far more improvement on average than other alternatives. If adaptive OFAT is enhanced by a well informed estimate of the preferred starting point design, its performance across every model variant is superior to every alternative method requiring similar number of runs and is comparable to alternatives requiring over four times as many runs.

In the second case study, the compound noise method has been analyzed on a theoretical basis to identify conditions under which it can reveal an optimum. The computational approach shows that even when the conditions for obtaining the optimum do not hold, the expected improvements may still be better than the alternatives (at least under some conditions).

A significant benefit of the computational approach presented here is that it allows robust design methods to be compared directly on the basis of their efficacy. At present, the choice among alternative robust design methods is often made on the basis of mathematical properties of the designs which only roughly correspond with effectiveness in practice. Alternatively, many engineers defend their robust design practices on the basis of past successes (this is often the case with Taguchi methods). It is hoped that this computational approach can be applied to offer more substantive evidence and better basis for method selection.

In practice, engineering designs are never truly optimal. Choices must be made to balance improvements in performance against time to market, cost, and other factors. The "80/20 rule" rings true to most practicing engineers who, day to day, apply limited resources across multiple activities in an uncertain environment. It is hoped that the computational approach presented here will eventually enable practitioners to make the difficult choices that arise deploying robust design in a realistic setting.

Chapter 5

Asymmetric Synergistic Interaction Structure

5.1 Overview

A new regularity is introduced and shown to be statistically significant in this chapter. It is shown that a preponderance of active two-factor interaction effects are synergistic, meaning that when main effects are used to increase the system response, the interaction provides an additional increase; and that when main effects are used to decrease the response, the interactions generally counteract the main effects. We name this newly observed system regularity Asymmetric Synergistic Interaction Structure (ASIS).

This regularity is observed and quantified by analyzing 113 published data sets which are described in details in Chapter 3. Based on a meta-analysis of the data in the form in which it is originally published, it is shown that about 80% of active interactions between two factors are synergistic. An illustrative example is presented to explain this new regularity. A further investigation of data transformation finds that ASIS is subject to change when data transformations are applied. My research shows that the log transformation tends to weaken ASIS or sometimes even reverse the structure.

5.2 Define ASIS

We use the term "asymmetric synergistic interaction structure" (ASIS) to describe the degree to which the signs of main effects provide information about the likely signs of interaction effects. Given the GLM described in Equation 1 or 1a, a synergistic two-factor interaction satisfies the inequality $\beta_i \beta_j \beta_{ij} > 0$ and an anti-synergistic two-factor interaction satisfies the inequality $\beta_i \beta_j \beta_{ij} < 0$.

Synergistic interactions do not always mean to enhance the effects of main factors. This can be easily shown with analysis on the sample scenarios. Consider a system with only two main effects and a two factor interaction. The all possible scenarios are listed in Table 30.

Table 23 A sample experiment for ASIS

Scenarios	β_1	β_2	β_{12}	Synergistic? (Y/N)	Examples
1	+	+	+	Yes	$E(y) = 2x_1 + 3x_2 + x_1 x_2$
2	+	+	-	No	$E(y) = 2x_1 + 3x_2 - x_1 x_2$
3	+	-	+	No	$E(y) = 2x_1 - 3x_2 + x_1 x_2$
4	-	+	-	Yes	$E(y) = -2x_1 + 3x_2 - x_1 x_2$
5	-	+	+	No	$E(y) = -2x_1 + 3x_2 + x_1 x_2$
6	+	-	-	Yes	$E(y) = 2x_1 - 3x_2 - x_1 x_2$
7	-	-	+	Yes	$E(y) = -2x_1 - 3x_2 + x_1 x_2$
8	-	-	-	No	$E(y) = -2x_1 - 3x_2 - x_1 x_2$

When the experimental responses are the-larger-the-better, for Scenario 1, we will set x_1 = +1 and x_2 = +1. Then, E(y) = 2*1 + 3*1 + 1. The synergistic interaction is helping to make improvement to the responses. However, when the experimental responses are the-smaller-the-better, also for Scenario 1, we will set x_1 = -1 and x_2 = -1. Then, E(y) = 2*(-1) + 3(-1) + 1*(-1)*(-1) = (-2) + (-3) + 1. The synergistic interaction is counteracting to main effects and impeding improvement to the responses. Similar analysis can be applied to other 7 scenarios.

With the defined *synergistic interactions* and *anti-synergistic interactions*, we define *asymmetric synergistic interaction structure* to be an effect relationship that the synergistic interactions and anti-synergistic interactions are not equally represented in a system. As our further analysis shows that synergistic interactions actually dominate in systems, ASIS will be specified to refer to this imbalance effect structure exist in engineering systems.

5.3 Quantifying ASIS

5.3.1 The Set of Experimental Data

We build a collection of 46 full factorial 2^k experiments and use their 113 data sets for the quantifying analysis. The data sets have been described in Section 3.3. A detailed list of all these experiments can be found in Appendix II. A complete list of the references where the original experimental data are presented can be found in Appendix I. The data sets include 569 two-factor interactions in which 63 two-factor interactions are cataloged as active effects using the Lenth method [39].

5.3.2 Method for Quantifying Asymmetric Synergistic Interaction Structure

To evaluate the null hypothesis that synergistic two-factor interactions and anti-synergistic two-factor interactions are equally likely, we followed these steps:

1) For each response
    a. Estimate the main effects and interactions for each response as described in Section 3.2.
    b. Label each two factor interaction as either synergistic or anti-synergistic according to our definition.
2) Carry out statistics on the set of 113 responses.
    a. Calculate the percentage of all two factor interactions that are synergistic and anti-synergistic.
    b. Use Lenth method to discriminate between active effects and inactive effects.
    c. Calculate the percentage of active two factor interactions that are synergistic and anti-synergistic.
    d. Calculate the percentage of inactive two factor interactions that are synergistic and anti-synergistic.
    e. Calculate 95% confidence intervals for the synergistic and anti-synergistic percentages using the binomial distribution.

5.4 Wet-Clutch Example Revisited

In Chapter 3, we used the wet-clutch experiment as an example to illustrate how we quantify effect spasity, hierarchy, and heredity regularities. We would also use this example here to illustrate our work in quantifying ASIS in this chapter.

In the wet-clutch experiment [41], there are seven factors under study. They are listed in Table 24.

Table 24 The factors in the wet clutch experiment

Factors	**Description**
A	Oil flow
B	Pack clearance
C	Spacer plate flatness
D	Friction material grooving
E	Oil viscosity
F	Friction material
G	Rotation speed

Following the procedure described in Section 5.3.2, we analyze the experimental data. The active main effects analysis has already been shown in Section 3.4. But for convenience, we re put the results into Table 25. Furthermore, we list the synergistic interaction analysis result here. The active two factor interactions as determined by the Lenth method are presented in Table 26 with the last column indicates whether it is synergistic or not.

Table 25 The main effects from the clutch case study.

Effect	Drag torque [ft lbs]	Active?
A	1.33	Yes
B	-1.55	Yes
C	-1.81	Yes
D	0.067	No
E	2.81	Yes
F	-0.092	No
G	3.01	Yes

Table 26 The active two factor interactions from the clutch case study.

Effect	Drag torque [ft lbs]	Synergistic?
AD	0.530	Yes
AG	0.964	Yes
BD	-0.520	Yes
BG	-0.830	Yes
CD	0.683	No
CG	-0.695	Yes
DE	0.642	Yes
DG	-0.914	No
EG	1.31	Yes

From the analysis, the hypothesized regularity, Asymmetric Synergistic Interaction Structure (ASIS) is strongly indicated. Seven of nine active two factor interactions meet the criterion because the sign of the interaction effect equals the sign of the product of the participating main effects.

This example raises an important point about ASIS. Many find the regularity to be surprising because, in their experience, a response becomes increasingly difficult to further improve as successive improvements are made. ASIS is not necessarily inconsistent with this general trend. In this example, to reduce drag torque, the main effects suggest that both oil flow

(*A*) and grooving (*D*) should be set to coded levels of -1. However, the significant *AD* interaction would lead to far less reduction of drag torque than one would expect from the linear model. In fact, the interactions will most likely determine the preferred level of *D* rather than the main effect.

To illustrate that non-linear transformation of responses can strongly affect regularities in data, especially ASIS, we applied a log transformation to the drag torque of the wet clutch pack experiment. All of our analysis is repeated. There is no difference in the active main effect analysis. However, the ASIS analysis of active two-factor interactions changed. The results are shown in Table 27.

Table 27 The active two factor interactions from the clutch case study using a log transform.

Effect	Log(Drag torque)	Synergistic?
AD	0.094	Yes
AG	0.159	Yes
BC	-0.072	No
BD	-0.096	Yes
BE	0.108	No
BG	-0.143	Yes
CD	0.182	No
CE	0.071	No
CF	-0.063	No
DE	0.103	Yes
DG	-0.228	No
EG	0.167	Yes

For this particular data set, after the log transformation, the number of active two factor interactions actually increased from 9 to 12. It is also important to note that in the original data, the synergistic interactions were more numerous, and in the transformed data the synergistic and anti-synergistic interactions are equally represented. This motivated an effort to assess the influence of transformations on ASIS through a second meta-analysis reported in Section 5.6.

5.5 Results of Meta-Analysis on ASIS

The methods described in Section 5.3 are applied to the set of 113 responses from published experiments. Table 28 presents the results of our investigation into Asymmetric Synergistic Interaction Structure (ASIS).

Table 28 Synergistic and anti-synergistic two-factor interactions in 113 experiments

		Synergistic	**Anti-Synergistic**	**Total**
All two-factor interactions	Number	362	207	569
	Percentage	64%	36%	100%
	Confidence interval (α=0.05)	60% to 68%	40% to 32%	
Active two-factor interactions	Number	52	11	63
	Percentage	83%	17%	100%
	Confidence interval (α=0.05)	71% to 91%	29% to 9%	

Firstly, it is noteworthy that about 2/3 of all two-factor interaction are synergistic. The confidence intervals for that percentage do not include 50%, so we can reject the null hypothesis that the two percentages might be equal.

Furthermore, it is of practical significance that the percentage of synergistic effects is much higher among active two-factor interactions than among all two-factor interactions.

5.6 Additional Investigation of the Log Transformation

The analysis in Section 5.5 is based on the data from experiments as originally published without any non-linear transformations. However, response transformations are common in analysis of experimental data. For background on good practice, see Wu and Hamada [6] who describe eight commonly used transformations. One motivation for transforming data is variance stabilization. Another is generation of a more parsimonious model with fewer higher order terms. To provide a rough sense of how such transformations affect the regularities reported here, we focused on just one commonly employed transformation, the logarithm.

We firstly analyzed which data sets were appropriate to apply the log transformation. We found that 107 out of 113 data sets could be subject to the log transformation as their responses contained only positive response values. Of the 107 data sets, it was found that log transformation resulted in more parsimonious models for 13 responses (meaning that the number of active effects were reduced), while the untransformed data produced more parsimonious models in 28 cases. In the other 66 responses, the number of significant effects was unaffected by the use of this transformation.

Reanalyzing those data sets after applied the log transformation, we observed that in both the full set of 107 transformed responses and in the smaller set of 13 more parsimonious transformed responses, the proportion of synergistic and anti-synergistic responses was not significantly different from 50%. An analysis of two factor interaction synergies on the log transformed data can be found in Table 29.

Table 29 Synergistic and anti-synergistic interactions in experiments whose responses were transformed using a logarithm.

			Synergistic	Anti-Synergistic	Total
Total 107 log-transformed data sets	All two-factor interactions	Number	271	268	539
		Percentage	50%	50%	100%
		C. I. (α=0.05)	46% to 55%	54% to 45%	
	Active two-factor interactions	Number	31	23	54
		Percentage	57%	43%	100%
		C. I. (α=0.05)	43% to 71%	57% to 29%	
13 data sets that became more parsimonious using the log-transform	All two-factor interactions	Number	37	41	78
		Percentage	47%	53%	100%
		C. I. (α=0.05)	36% to 59%	64% to 41%	
	Active two-factor interactions	Number	2	3	5
		Percentage	40%	60%	100%
		C. I. (α=0.05)	5% to 85%	95% to 15%	

Therefore, we conclude that the newly reported regularity of asymmetric synergistic interaction structure (ASIS) is a property of data as they are reported by their experimenters (usually in physical dimensions) and is not generally persistent under non-linear transformations of the reported data. ASIS is a function of the physical systems and whatever transformations experimenters actually use before reporting the data, but may be altered by further transformation.

5.7 Conclusions and Discussion

The major outcome of this study is identification and quantification of Asymmetric Synergistic Interaction Structure (ASIS) -- a strong regularity not previously identified in the literature. It was shown that about 80% of active two-factor interactions are synergistic, meaning that $\beta_i \beta_j \beta_{ij} > 0$. The consequences of ASIS for engineering design require further discussion.

In cases wherein larger responses are preferred, procedures that exploit main effects are likely to enjoy additional increases due to active two-factor interactions even if those interactions have not been located or estimated. By contrast, in cases wherein smaller responses are preferred, procedures that exploit main effects to reduce the response are likely to be penalized by increases due to active two-factor interactions.

The discussion of ASIS and its relationship to improvement efforts raises the question of why ASIS was defined as it was in this paper. This definition was chosen because it revealed the new, statistically significant regularity in the data set. Other relationships among main effects and interactions were explored and found to be insignificant. However, any regularity associated with improvements rather than increases raise practical and conceptual difficulties. This study was based on meta-analysis of published data sets. If the authors of published data sets do not clearly state whether larger or smaller responses are preferred, how can one define "improvement" for that data set? Further, even if the authors express a preference, might not a different application of the same physical phenomenon reverse that preference? By contrast, regularities associated with the published values reflect regularities in physical phenomena as observed and interpreted by the experimenters. To the extent that such regularities exist and can be confirmed as stable and reliable, they can be helpful in interpreting data.

Some experienced practitioners will find ASIS surprising. It is common for experimenters to report that, if they use experimentation to attain some increases in a response, then any further increase will be harder to attain. We agree that this is the general trend in engineering quality improvements, but how our proposed synergy concept relates to this issue is not so simple. When engineers seek to improve a system, they move toward regions of improvement until locating local maxima or constraints. These maxima and constraints make additional improvements difficult to achieve. Our results are based on meta-analysis of 2^k experiments. It is an interesting question whether such experiments are typically conducted at local maxima or away from them. If 2^k experiments are typically conducted away from local maxima, there are at least two explanations:

1) The maximum has not yet been located;
2) Constraints on the design space are limiting the optimization of that engineering system.

Determining the underlying reasons for ASIS is an interesting subject for future research. It is odd that such a strong regularity has not been discussed in either theoretical or practical discourse regarding design of experiments (DOE). The previously known regularities of effect sparsity, hierarchy, and heredity are intellectual cornerstones of DOE and many popular methods provide benefit by exploiting them. Perhaps future research will give rise to new DOE methods that exploit ASIS and thereby reduce resource demands and/or increase effectiveness of engineering experimentation.

One example of potential applications of ASIS is in Robust Design. There has been substantial discussion and analysis of "compound noise" as an approach to reducing resource requirements for robust design. Hou et. al. [56] argued that compound noise is unlikely to be effective since so-called “extreme conditions” used to form compound noises are unlikely to be consistent as control factors are varied. On the other hand, many practitioners argue that compound noise is usually effective. It seems possible that the results presented here can resolve this issue and bring theory and practice back into better alignment.

Another example lays in Bayesian model selection methods in design of experiments. Incorporating ASIS into the Bayesian model might lead us to a much more accurate description of subject systems comparing to set equal possibilities to synergistic and anti-synergistic interactions. We expect a large improvement associated with models incorporating ASIS.

5.8 Summary

This chapter introduced a new regularity which we called *Asymmetric Synergistic Interaction Structure* or *ASIS*. It was identified and validated that a preponderance of active two-factor interaction effects is synergistic. A quantitative study of ASIS was presented and statistics on ASIS over our data sets was reported. It seems surprising that this regularity exists and is so significant (It was shown that about 80% of active two-factor interactions are synergistic). The mechanisms giving rise to this regularity are not known and might be a good subject for further research. This study of ASIS should enable development of more efficient experimental plans for response surface methods, robust design, and other engineering methodologies.

Chapter 6

Conclusions

6.1 Major Contributions

My research contributions can be summarized into four major aspects.

(1) Build A Large Engineering Experimental Database.

An engineering experimental database is built-up with 113 full factorial engineering experimental data sets. All data are collected from a wide range of published engineering experiments with the number of factors in those experiments varying from 3 to 7. The database has several valuable features:

- High Resolution – All data sets are full factorial experiments which offer sufficient degree of freedom to analyze both main effects and interactions in each experiment;
- Reliable Data Source – All data are collected from published experiments which are peer reviewed;
- Representative Systems – All experiments are carried out on engineering systems from a variety of fields. They represent general engineering experimental practice and offer intuitions and information for a general system;
- Variety of Factors – The data sets include a range of factors from 3 to 7 with the majority of experiments contain 3 or 4 factors.

The database can either be used in DOE theoretical analysis and new method development or be analyzed to offer empirical parameters for experimental practitioners.

(2) Verify and Quantify Three System Regularities.

For the first time, the three system regularities discussed in the DOE community, i.e., effect sparsity, hierarchy, and heredity, are verified and quantified. These three regularities could essentially affect the effectiveness of DOE and of Robust Design methods. Results show strong evidence that these regularities exist in engineering systems but also suggest certain caveats when applying them in experimental analysis.

(3) Develop a New Model-Based Strategy for Validating and Comparing Robust Design Methods.

A hierarchical probability model is developed to generate general engineering system responses. Using this model, a new strategy is created to compare the effectiveness of different robust design methods. This strategy fills the gap between pure theoretical analysis and practical experience. It can be used in new robust design development, robust design method validation and comparison, and robust design effectiveness estimation.

An adaptive approach is introduced into robust designs for the first time. In the case study, it is shown that using adaptive OFAT crossed with an orthogonal outer array results in significantly more improvement, on average, than other alternatives. If adaptive OFAT is enhanced by a well informed estimate of the preferred starting point design, its performance across every model variant is superior to every alternative method requiring similar number of runs and is comparable to alternatives requiring over four times as many runs.

(4) Establish a New System Regularity – ASIS.

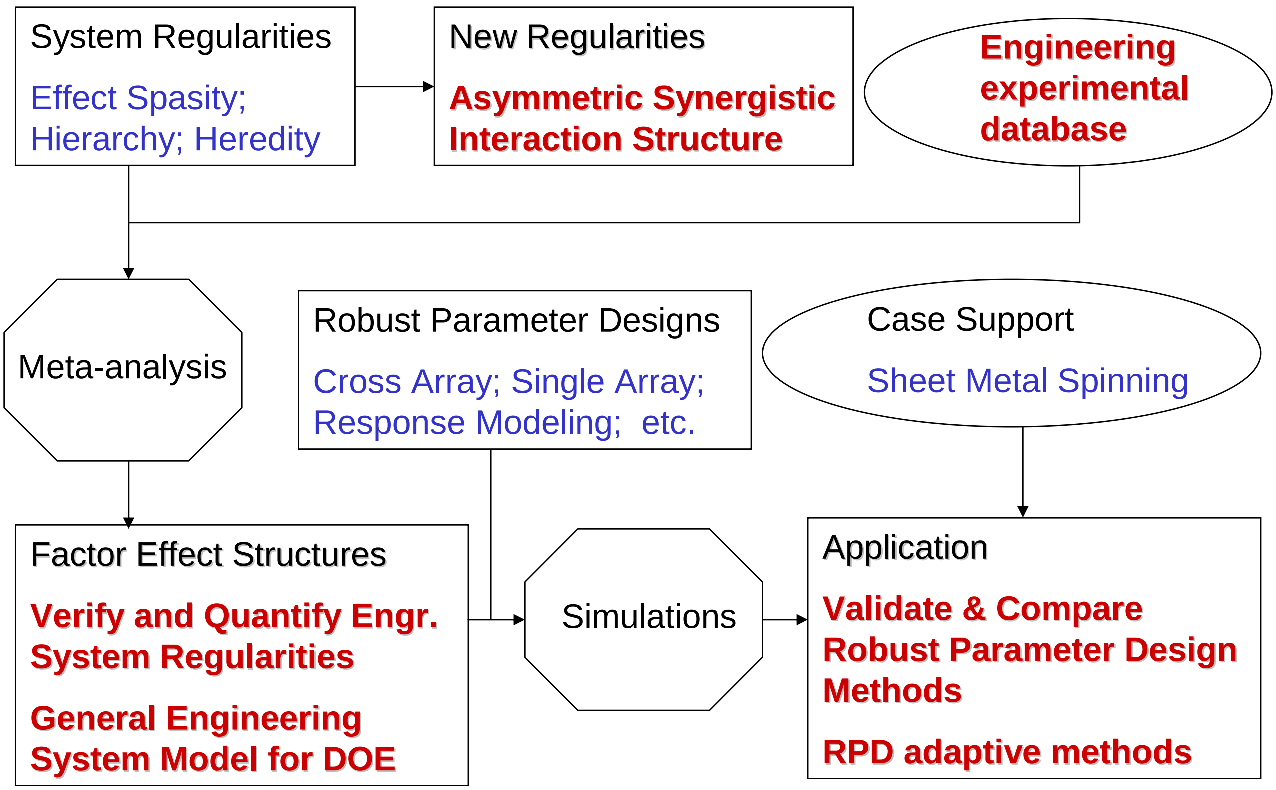
A new system regularity which we named Asymmetric Synergistic Interaction Structure or ASIS is found and quantified from our database analysis. It is shown that that about 80% of active interactions between two factors are synergistic, meaning that when main effects are used to increase the system response, the interaction provides an additional increase. It is asymmetric that when main effects are used to decrease the response the interactions generally counteract the main effects. The mechanisms giving rise to this regularity are not known and might be a good subject for further research. This study of ASIS should enable development of more efficient experimental plans for response surface methods, robust design, and other engineering methodologies.

6.2 Summary of My Work

My work can be summarized with the flowchart shown in Figure 20.

I firstly built the database. Then I used the database to verify and quantify three system regularities and found a new regularities ASIS. Incorporating the system regularities into the hierarchical probability model enabled me to develop a new strategy for the validation and comparison of robust design methods. My results support the observation from the sheet metal spinning experiment [22] and the new strategy answered all questions discussed in the motivation section of this dissertation (Section 1.2).

Figure 20 Summary of my work



6.3 Future Work

This work inspired several directions for future research. One of them is to expand the engineering experimental database. To aggregate more experiments into the database will enable a further categorization of those engineering systems, either by system regularities, or by fields of engineering practice, or by number of factors. Based on the results, system regularities and other features can be re-analyzed with the benefit to compare different categories and understand the relationship between them. This would offer higher resolution results of those discussed in this dissertation and could provide more detailed instruction to practitioners according to their system characteristics.

Another major direction is to carry out an advanced study on adaptive robust design methods. In the case study of this work, it is shown that introducing adaptive methods into robust design greatly enhanced the performance. A further analysis can follow with theoretical reasoning and case studies to support this approach. It may result in a revolutionary method which could greatly enhance robust design performance on a large number of systems.

The third promising direction is to further investigate ASIS. To incorporate ASIS into the probability model will certainly affect setting up priors for the Bayesian model selection approach. New DOE or Robust Design methods might be created based on this new regularity.

Appendix I. References of the Experimental Data Sets

1. Admassu, W.; Breese, T. Feasibility of using natural fishbone apatite as a substitute for hydroxyapatite in remediating aqueous heavy metals. Journal of Hazardous Materials 1999, B69; pp 187-196.
2. Bergman, R.S.; Cox, C.W.; DePriest, D.J.; Faltin, F.W. Effect of process variations on incandescent lamp performance. Journal of the Illumination Engineering Society 1990, 19-2; pp 132-141.
3. Bogoeva-Gaceva, G.; Mader, E.; Queck, H. Properties of glass fiber polypropylene composites produced from split-warp-knit textile preforms. Journal of Thermoplastic Composite Materials 2000, 13; pp 363-377.
4. Brachet, A.; Rudaz S.; Mateus, L.; Christen, P.; Veuthey, J.L. Optimization of accelerated solvent extraction of cocaine and benzolecgonine from coca leaves. Journal of Separation Science 2001, 24; pp 865-873.
5. Friedman, M.; Savage, L. J. Planning experiments seeking maxima. Techniques of Statistical Analysis, 1947; pp 365-372.
6. Grimm, J.; Chlebek, J.; Schulz, T.; Huber, H.L. The influence of post-exposure bake on line width control for the resist system RAY-PN (AZ PN 100) in x-ray mask fabrication. J. Vac. Sci. Technol. 1991, B9-6; pp 3392-3398.
7. Karthikeyan, R.; Lakshmi Narayanan, P.R.; Naagarazan, R.S. Mathematical modeling for electric discharge machining of aluminum-silicon carbide particulate composites. Journal of Materials Processing Technology 1999, 87; pp 59-63.

8. Laus, M.; Lelli, M.; Casagrande, A. Polyepichlorohydrine stabilized core shell microspheres by dispersion polymerization. Journal of Polymer Science, Part A: Polymer Chemistry 1997, 35-4; pp 681-688.
9. Lee, S.S.G.; Tam, S.C.; Loh, N.H.; Miyazawa, S. An investigation into the ball burnishing of an ANSI 1045 free-form surface. Journal of Materials Processing Technology 1992, 29; pp 203-211.
10. Modi, O.P.; Yadav, R.P.; Mondal, D.P.; Dasgupta, R.; Das, S.; Yegneswaran, A.H. Abrasive wear behavior of zinc-aluminum alloy – 10% Al_2O_3 composite through factorial design of experiment. Journal of Materials Science 2001, 36; pp 1601-1607.
11. Moskowitz, I. L.; Babu, S.V. Surface morphology and quality of a-Si:C:H films. Thin Solid Films 2001, 385; pp 45-54.
12. Murugan, N.; Parmar, R.S. Effects of MIG process parameters on the geometry of the bead in the automatic surfacing of stainless steel. Journal of Materials Processing Technology 1994, 41-4; pp 381-398.
13. Myers, R.H.; Montgomery, D.C. Response Surface Methodology. John Wiley & Sons: New York, 1995, p. 105.
14. Olofsson, U.; Holmgren, M. Friction measurement at low sliding speed using a servohydraulic tension-torsion machine. Experimental Mechanics 1994, 34-3; pp 202-207.
15. Saravanan, P.; Selvarajan, V.; Joshi, S.V.; Sundararajan, G. Experimental design and performance analysis of alumina coatings deposited by a detonation spray process. Journal of Physics D: Applied Physics 2001, 34; pp 131-140.
16. Sen, R. Response surface optimization of the critical media components for the production of surfactin. J. Chem. Tech. Biotechnol. 1997, 68; pp 263-270.

17. Sharma, A.K.; Forester, W.K.; Shriver, E.H. Physical and optical properties of steam-exploded laser-printed paper. TAPPI Journal 1996, 79-5; pp 211-221.
18. Shearer, G.; Tzoganakis, C. Free radical hydrosilylation of polypropylene. Journal of Applied Polymer Science 1996, 65-3; pp 439-447.
19. Shulka, A.K.; Stevens, P.; Hamnett, A.; Goodenough, J.P. A nafion-bound platinized carbon electrode for oxygen reduction in solid polymer electrolyte cells. Journal of Applied Electrochemistry 1989, 19; pp 383-386.
20. Smith, S.D.; Osbourne, J.R.; Forde, M.C. Analysis of earth moving systems using discrete event simulation. Journal of Construction Engineering and Management 1995, 121-4; pp 388-396.
21. Sosada, M. Optimal conditions for fractionation of rapeseed lecithin with alcohols. JAOCS 1993, 70-4; pp 405-410.
22. Thompson, N.G.; Islam, M.; Lankard, D.A.; Virmani, Y.P. Environmental factors in the deterioration of reinforced concrete. Materials Performance 1995, 34-9; pp 43- 47.
23. Wang, X.; Feng, C.X.; He, D.W. Regression analysis and neural networks applied to surface roughness study in turning. IIE Transactions on Design and Manufacturing 2002.
24. Wu, C.F.J.; Hamada, M. Experiments: Planning, Design, and Parameter Optimization. John Wiley & Sons, Inc.: New York, 2000
25. Vuk, T.; Tinta, V.; Gabrovsek, R.; Kaucic, V. The effect of limestone addition, clinker type and fineness on properties of Portland cement. Cement and Concrete Research 2001, 31; pp 135-139.
26. Dirilgen, N. Effects of pH and CHELATOR EDTA on Cr TOXICITY and ACCUMULATION in LEMNA minor, Chemosphere 1998, 37-4; pp 771-783.

27. Taylor, J.B.; Carrano, A.L.; Lemaster, R.L. Quantification of process parameters in a wood sanding operation. Forest Products Journal 1999, 49-5; pp 41-46.
28. Pei, Z.J.; Xin, X.J.; Liu, W. Finite element analysis for grinding of wire-sawn silicon wafers: a designed experiment. International Journal of Machine Tools & Manufacture 2003, 43; pp 7-16.
29. Soudki, K.A.; El-Salakawy, E.F.; Elkum, N.B. Full factorial optimization of concrete mix design for hot climates. Journal of Materials in Civil Engineering, Nov/Dec, 2001; pp 427-433.
30. Brabham, D.E. Designing color performance into color-improved HPS lamps. Journal of the Illuminating Engineering Society, winter, 1991; pp 50-55
31. Rahman, M.; Ramakrishna, S.; Thoo, H.C. Machinability study of carbon/peek composites. Machining Science and Technology 1998, 3-1; pp 49-59.
32. Dini, J.W.; Kelley, W.K.; Cowden, W.C.; Lopez, E.M. Use of electrodeposited silver as an aid in diffusion welding. Welding Research Supplenment, January, 1984; pp 26s-34s.
33. Gurses, A.; Yalcin, M.; Dogar, C. Electrocoagulation of some reactive dyes: a statistical investigation of some electrochemical variables. Waste Management 2002, 22; pp 491-499.
34. Pei, Z.J.; Strasbaugh, A. Fine grinding of silicon wafers: designed experiments. International Journal of Machine Tools & Manufacture 2002, 42; pp 395-404.
35. Sahoo, R.N.; Naik, P.K.; Das, S.C. Leaching of manganese from low-grade manganese ore using oxalic acid as reductant in sulphuric acid solution. Hydrometallurgy 2001, 62; pp 157-163.

36. Naik, P.K.; Sukla, L.B.; Das, S.C. Aqueous SO2 leaching studies on Nishikhal manganese ore through factorial experiment. Hydrometallurgy 2000, 54; pp 217-228.

37. Schultz, M.K.; Inn, K.G.W.; Lin, Z.C.; Burnett, W.C.; Smith, G.; Biegalski, S.R.; Filliben, J. Identification of radionuclide partitioning in soils and sediments: determination of optimum conditions for the exchangeable fraction of the NIST standard sequential extraction protocol. Appl. Radiat. Isot. 1998, 49 (9-11); pp 1289-1293.

38. Carter Jr, C.W.; Doublie, S.; Coleman, D.E. Quantitative analysis of crystal growth: Tryptophanyl-tRNA synthetase crystal polymorphism and its relationship to catalysis. J. Mol. Biol. 1994, 238; pp 346-365.

39. Becerra, M.; Gonzalez Siso, M.I. Yeast b-galactosidase in solid-state fermentations. Enzyme and Microbial Technology 1996, 19; pp 39-44.

40. Asanza Teruel, M.L.; Gontier, E.; Bienaime, C.; Nava Saucedo, J.E.; Barbotin, J.-N. Response surface analysis of chlortetracycline and tetracycline production with K-carrageenan immobilized streptomyces aureofaciens. Enzyme and Microbial Technology 1997, 21; pp 314-320.

41. Trezona, R.I.; Pickles, M.J.; Hutchings, I.M. A full factorial investigation of the erosion durability of automotive clearcoats. Tribology International 2000, 33; pp 559-571.

42. Gupte, M.; Kulkarni, P. A study of antifungal antibiotic production by Thermomonospora sp MTCC 3340 using full factorial design. Journal of Chemical Technology and Biotechnology 2003, 78; pp 605-610.

43. Mussatto S.I.; Roberto, I.C. Optimal experimental condition for hemicellulosic hydrolyzate treatment with activated charcoal for xylitol production. Biotechnol. Prog. 2004, 20; pp 134-139.

44. Pan, T.-Y.; Cooper, R.R.; Blair, H.D.; Whalen, T.J.; Nicholson, J.M. Experimental analysis of thermal cycling fatigue of four-layered FR4 printed wiring boards. Journal of Electronic Packaging 1994, 116; pp 76-78.

45. Spedding, T.A.; Wang, Z.Q. Study on modeling of wire EDM process. Journal of Materials Processing Technology 1997, 69 (1-3); pp 18-28.

46. Lloyd, F.A. Parameters contributing to power loss in disengaged wet clutches. Society of Automotive Engineers Transactions 1974, 83; pp 2498-2507.

Appendix II. List of the Responses Subjected to Meta-Analysis

Engineering System [Ref. #]§	k^n	Response	Engineering System [Ref. #]	k^n	Response
Remediating aqueous heavy metals [1]	2^4	Lead	Finish turning [23]	2^5	roughness
Remediating aqueous heavy metals [1]	2^4	pH	Epitaxial layer growth [24]	2^4	thickness
Remediating aqueous heavy metals [1]	2^4	Lead	Limestone Effects [25]	2^3	Surface Area
Remediating aqueous heavy metals [1]	2^4	pH (alt. method)	Limestone Effects [25]	2^3	Water Demand
Processing of incandescent lamps [2]	2^4	lumens fluct.	Limestone Effects [25]	2^3	Init. Setting Time
Processing of incandescent lamps [2]	2^4	power val.	Limestone Effects [25]	2^3	Final Setting Time
Processing of incandescent lamps [2]	2^4	Lumens val.	Cr Toxicity & L. nimor [26]	2^{3*}	RGR
Processing of incandescent lamps [2]	2^4	life time	Cr Toxicity & L. nimor [26]	2^{3*}	DFR
Processing of incandescent lamps [2]	2^4	power fluct.	Cr Toxicity & L. nimor [26]	2^{3*}	Cr In Fronds
Processing of incandescent lamps [2]	2^4	life fluct.	Wood Sanding Oper. [27]	2^{4*}	Cherry Removal Rate
Glass fiber composites [3]	2^3	stiffness tans.	Wood Sanding Oper. [27]	2^{4*}	Maple Removal Rate
Glass fiber composites [3]	2^3	Stiffness	Wood Sanding Oper. [27]	2^{4*}	Oak Removal Rate
Glass fiber composites [3]	2^3	strength	Wood Sanding Oper. [27]	2^{4*}	Pine Removal Rate
Glass fiber composites [3]	2^3	strength trans.	Wood Sanding Oper. [27]	2^{4*}	Cherry Surface Rough
Solvent extraction of cocaine [4]	2^3	% weight	Wood Sanding Oper. [27]	2^{4*}	Maple Surface Rough
Plasma spraying of ZrO2 [5]	2^4	velocity	Wood Sanding Oper. [27]	2^{4*}	Oak Surface Rough
Plasma spraying of ZrO2 [5]	2^4	Temp.	Wood Sanding Oper. [27]	2^{4*}	Pine Surface Rough
Plasma spraying of ZrO2 [5]	2^4	Size	Grinding of silicon wafers[28]	2^4	Displacement
Post-exp. bake in x-ray mask fab. [6]	2^4	line width	Concrete Mix Hot Clim. [29]	2^{4*}	Compressive Strength
EDM of carbide composites [7]	2^{3*}	roughness	Color-Improved Lamps [30]	2^4	Voltage
EDM of carbide composites [7]	2^{3*}	tool wear	Color-Improved Lamps [30]	2^4	CCT
EDM of carbide composites [7]	2^{3*}	MRR	Color-Improved Lamps [30]	2^4	CRI
Polymerization of microspheres [8]	2^3	M_n	Color-Improved Lamps [30]	2^4	Luminous Flux
Polymerization of microspheres [8]	2^3	surf. density	Machinability Study [31]	2^4	Tool Wear
Polymerization of microspheres [8]	2^3	diameter	Machinability Study [31]	2^4	Surface Finish
Polymerization of microspheres [8]	2^3	M_w	Diffusion Welding [32]	2^4	Failure Load
Polymerization of microspheres [8]	2^3	M_z	Electrocoagulation [33]	2^4	Decolorization
Polymerization of microspheres [8]	2^3	surface density	Fine grinding [34]	2^3	Max grinding force
Polymerization of microspheres [8]	2^3	% pepi	Fine grinding [34]	2^3	Max motor current
Ball burnishing of an ANSI 1045 [9]	2^4	roughness	Fine grinding [34]	2^3	Grinding cycle time
Abrasive wear of Zi-Al alloy [10]	2^3	Zinc	Fine grinding [34]	2^3	Surface roughness
Abrasive wear of Zi-Al alloy [10]	2^3	composite	Leaching of Manganese [35]	2^4	Mn
Surface morphology of films [11]	2^5	roughness	Leaching of Manganese [35]	2^4	Fe
Surface morphology of films [11]	2^5	stress	Leaching of Manganese [35]	2^4	Al
MIG process [12]	2^4	penetration	Aqueous SO2 Leaching [36]	2^4	Extraction Mn
MIG process [12]	2^4	reinforce	Aqueous SO2 Leaching [36]	2^4	Extraction Fe
MIG process [12]	2^4	width	Ident. of Radionuclide [37]	2^3	U^{238} extracted
MIG process [12]	2^4	reinforce	Crystal Growth [38]	2^4	Experimental scores
Pilot plant filtration rate [13]	2^4	Rate	Yeast b-G [39]	2^3	Observed b-G
Friction measurement machine [14]	2^3	frict coeff val.	Chl & Tetracycline [40]	2^{3*}	CTC
Friction measurement machine [14]	2^3	frict coeff fluct.	Chl & Tetracycline [40]	2^{3*}	TC
Detonation spray process [15]	2^4	hardness	Chl & Tetracycline [40]	2^{3*}	pH
Detonation spray process [15]	2^4	roughness	Erosion Durability [41]	2^{3*}	Nozzle Pressure
Detonation spray process [15]	2^4	porosity	Antifungal antibiotic [42]	2^{3*}	Antifungal Antibio. Act.
Production of surfactin [16]	2^4	Yield	Xylitol Production [43]	2^4	CR
Steam-exp. laser-printed paper [17]	2^5	brightness	Xylitol Production [43]	2^4	LDPR
Steam-exp. laser-printed paper [17]	2^5	opacity	Xylitol Production [43]	2^4	Yp/s
Steam-exp. laser-printed paper [17]	2^5	light abs.	Xylitol Production [43]	2^4	Qp
Steam-exp. laser-printed paper [17]	2^5	light scatter	Thermal fatigue of PWBs [44]	2^3	cycles
Hydrosilylation of polypropylene [18]	2^3	silane	Wire EDM process [45]	2^3	Roughness (μ)
Hydrosilylation of polypropylene [18]	2^3	Double	Wire EDM process [45]	2^3	waviness (μ)
Solid polymer electrolyte cells [19]	2^3	potential	Wire EDM process [45]	2^3	cut spd (μ)
Simulation of earth moving sys. [20]	2^6	match factor	Wire EDM process [45]	2^3	Roughness (σ)
Simulation of earth moving sys. [20]	2^6	production	Wire EDM process [45]	2^3	cut spd (σ)
Fractionation of rapeseed lecithin [21]	2^4	enrichment	Wire EDM process [45]	2^3	waviness (σ)
Fractionation of rapeseed lecithin [21]	2^4	Yield	Wet clutch pack [46]	2^7	drag torque
Deter. of reinforced concrete [22]	2^{3*}	corros. rate			

** Only the full factorial data subsets in these experiments were used in the meta-analysis in this dissertation.*

§ *The reference number matches the reference list in Appendix I.*

Bibliography

[1] Thomke, S., 2001, Enlightened Experimentation: The New Imperative for Innovation. Harvard Business Review; pp. 67-75.

[2] Frey, D. D., and Li, X., 2005, Using Hierarchical Probability Models to Evaluate Robust Parameter Design Methods, submitted to Journal of Quality Technology.

[3] Box, G. E. P., Hunter, W. G., and Hunter, J. S., 1978, Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, John Wiley & Sons Inc., New York.

[4] Logothetis, N.; Wynn, H. P., 1994, Quality Through Design: Experimental Design, Off-line Quality Control and Taguchi's Contributions, Clarendon Press, Oxford.

[5] Czitrom, V., 1999, One-Factor-at-A-Time versus Designed Experiments, The American Statistician, 53, pp. 126-131.

[6] Wu, C. F. J., and Hamada, M., 2000, Experiments: Planning, Analysis, and Parameter Design Optimization, Wiley & Sons Inc., New York.

[7] Friedman, M., and Savage, L. J., 1947, Planning Experiments Seeking Maxima, Techniques of Statistical Analysis, pp. 365-372.

[8] Daniel, C., 1973, One-at-a-Time Plans, Journal of the American Statistical Association, 68, pp. 353-360.

[9] Koita, R., 1994, Strategies for Sequential Design of Experiments, SM Thesis, Massachusetts Institute of Technology.

[10] McDaniel, W. R., and Ankenman, B. E., 2000, Comparing Experimental Design Strategies for Qualigy Improvement with Minimal Changes to Factor Levels, Quality and Reliability Engineering International, 16, pp. 355-362.

[11] Frey, D. D., Engelhardt, F., and Greitzer, E. M., 2003, A Role for One-Factor-at-a-Time Experimentation in Parameter Design, Research in Engineering Design, 14(2), pp. 65-74.

[12] Taguchi, G., and Wu, Y., 1985, Introduction to Off-line Quality Control, Central Japan Quality Control Association, Tokyo.

[13] Kacker, R. N., 1985, Off Line Quality Control, Parameter Design, and the Taguchi Method, Journal of Quality Technology, 17, pp. 176-188.

[14] Hunter, J. S., 1985, Statistical Design Applied to Product Design, Journal of Quality Technology, 17, pp. 210-221.

[15] Box, G. E. P., 1985, Discussion: Off Line Quality Control, Parameter Design, and the Taguchi Method, Journal of Quality Technology, 17, pp. 189-190.

[16] Box, G. E. P., 1988, Signal-to-Noise Ratios, Performance Criteria, and Transformations (with discussion), Technometrics, 30, pp. 1-40.

[17] Hamada, M., and Wu, C. F. J., 1990, A Critical Look at Accumulation Analysis and Related Methods (with discussion), Technometrics, 32, pp. 119-130.

[18] Nair, V. N., Editor, 1992, Taguchi's Parameter Design: A Panel Discussion, Technometrics, 34, pp. 127-161.

[19] 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, pp. 15-22.

[20] Shoemaker, A. C., Tsui, K. L., and Wu, C. F. J., 1991, Economical Experimentation Methods for Robust Design, Technometrics, 33, pp. 415-427.

[21] Borror, C. M., and Montgomery, D. C., 2000, Mixed Resolution Designs as Alternatives to Taguchi Inner/Outer Array Designs for Robust Design Problems, Quality and Reliability International, 16, pp. 117-127.

[22] Kunert, J., Auer, C., Erdbrugge, M., and Gobel, R., 2005, An Experiment to Compare the Combined Array and the Product Array for Robust Parameter Design, accepted to Journal of Quality Technology.

[23] Box, G. E. P., and Draper, N. R., 1987, Empirical Model-Building and Response Surface. John Wiley & Sons Inc. New York.

[24] Taguchi, G., translated by Tung, L. W., 1987, System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Costs. QualityResources: A Division of the Kraus Organization Limited, New York; American Supplier Institute, Inc., Dearborn, MI.

[25] Phadke, M. S., 1989, Quality Engineering Using Robust Design. PTR Prentice-Hall: A Simon & Schuster Company, Englewood Cliffs, NJ.

[26] Engineering Statistics Handbook; http://www.itl.nist.gov/div898/handbook/index.htm

[27] Fisher, R. A., 1949, The Design of Experiments (5th Edition), Oliver and Boyd, London and Edinburgh.

[28] Myer, R. H., and Montgomery, D. C., 1995, Response Surface Methodology: Process and Product Optimization Using Designed Experiments, John Wiley & Sons Inc., New York.

[29] Montgomery, D. C., 2004, Design and Analysis of Experiments, John Wiley & Sons Inc., New York.

[30] Frey, D. D., and Wang, H. J., 2006, Adaptive One-Factor-at-A-Time Experimentation and Expected Value of Improvement, accepted to Technometrics.

[31] Frey, D. D., 2003, How One-Factor-at-a-Time Experimentation Can Lead to Greater Improvements than Orthogonal Arrays, DETC 2003/DTM48646, ASME Design Engineering Technical Conference, September 2-6, 2003, Chicago, IL.

[32] Frey, D. D., and Li, X., 2004, Validating Robust Parameter Design Methods. ASME Design Engineering Technical Conference, DETC2004-57518, Salt Lake City, Utah.

[33] Box, G. E. P., and Meyer, R. D., 1986, An Analysis for Unreplicated Fractional Factorials, Technometrics, 28, pp. 11-18.

[34] McKay, M. D., Beckman, R. J., and Conover, W. J., 2000, Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics, 42(1), pp. 55-61.

[35] Hamada, M., and Wu, C. F. J., 1992, Analysis of Designed Experiments with Complex Aliasing, Journal of Quality Technology, 24, pp. 130-137.

[36] Box, G. E. P., and Liu, P. T. Y., 1999, Statistics as a Catalyst to Learning by Scientific Method, Journal of Quality Technology, 31(1), pp. 1-29.

[37] Chipman, H. M., Hamada, M., and Wu, C. F. J., 1997, Bayesian Variable-selection Approach for Analyzing Designed Experiments with Complex Aliasing. Technometrics, 39(4), pp. 372-381.

[38] Daniel, C., 1959, Use of Half-normal Plots in Interpreting Factorial Two-level Experiments, Technometrics, 1, pp. 311-341.

[39] Lenth, R. V., 1989, Quick and Easy Analysis of Unreplicated Factorials. Technometrics, 31(4), pp. 469-473.

[40] Sharma, A. K., Forester, W. K., and Shriver, E. H., 1996, Physical and Optical Properties of Steam-Exploded Laser-Printed Paper, Tappi Journal, 79(5), pp. 211-221.

[41] Lloyd, F. A., 1974, Parameters Contributing to Power Loss in Disengaged Wet Clutches, Society of Automotive Engineers Transactions, 83, pp. 2498-2507.

[42] Kunert, J., 2004, A Comparison of Taguchi's Product Array and the Combined Array in Robust-parameter-design, Spring Research Conference on Statistics in Industry and Technology, Gaithersburg, MD.

[43] American Institute of Aeronautics and Astronautics, 1998, Guide for the Verification and Validation of Computational Fluid Dynamics Simulations, AIAA G-077-1998.

[44] Hazelrigg, G.A., 2003, Thoughts on Model Validation for Engineering Design, DETC 2003/DTM48632, ASME Design Engineering Technical Conference, September 2-6, Chicago, IL.

[45] Institute of Electrical and Electronics Engineers, 1998, IEEE Standard for Software Verification and Validation, IEEE Std. 1012-1998.

[46] Olewnik, A.T., and Lewis, K.E., 2003, On Validating Design Decision Methodologies, DETC 2003/DTM48669, ASME Design Engineering Technical Conference, September 2-6, Chicago, IL.

[47] Wu, C. F. J, and Zhu, Y., 2003, Optimal Selection of Single Arrays for Parameter Design Experiments, Statistica Sinica, 13, pp. 1179-1199.

[48] Steinberg, D. M., and Bursztyn, D., 1994, Dispersion Effects in Robust-design Experiments with Noise Factors, Journal of Quality Technology, 26, pp. 12-20.

[49] Vining, G. G., and Meyers, R. H., 1990, Combining Taguchi and Response Surface Philosophies: A Dual Response Approach, Journal of Quality Technology, 22, pp. 38-45.

[50] Box, G. E. P., and Wilson, K. B., 1951, On the Experimental Attainment of Optimum Conditions, Journal of the Royal Statistical Society, B13, pp. 1-38.

[51] Prozanto, L., 2000, Adaptive Optimization and D-Optimum Experimental Design, Annals of Statistics, 28, pp. 1743-1761.

[52] Henkenjohann, N., Gobel, R., Kleiner, M., and Kunert, J., 2005, An Adaptive Sequential Procedure for Efficient Optimization of the Sheet Metal Spinning Process, Quality and Reliability Engineering International, 21, pp. 439-455.

[53] Li, X., and Frey, D. D., 2005, Regularities in Data from Factorial Experiments, accepted to Complexity.

[54] Morrison, S. J., 1957, Study of Variability in Engineering Design, Applied Statistics, 6, pp. 133 -138.

[55] Easterbrook. P. J., Berlin, J. A., Gopalan, R., and Mathews, D. R., 1991, Publication Bias in Clinical Research, Lancet, 337, pp. 867-872.

[56] Hou, X. S., 2002, On the Use of Compound Noise Factor in Parameter Design Experiments, Applied Stochastic Models in Business and Industry, 18, pp. 225-243.