THE PRINCIPLES OF DESIGN APPLIED TO ENGINEERING AND POLICY

by

STEPHEN FRANCIS FILIPPONE

B.S. Electrical Engineering, Northeastern University
(1985)

SUBMITTED TO THE DEPARTMENT OF
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

MASTER OF SCIENCE IN TECHNOLOGY AND POLICY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
August, 1988

Signature of Author ______________________________
Department of Electrical Engineering and Computer Science
August 5, 1988

Certified by ______________________________
Nam P. Suh
Professor of Mechanical Engineering
Thesis Supervisor

Accepted by ______________________________
Richard de Neufville
Professor of Civil Engineering
Chairman, Technology and Policy Program

This Thesis is Dedicated to My Mother

Joan Frances Filippone

and to My Grandmother

Sarah Frances Mullevey

For the twenty six years of sacrifice endured in the interest of my education.

THE PRINCIPLES OF DESIGN APPLIED TO ENGINEERING AND POLICY

by

STEPHEN FRANCIS FILIPPONE

Submitted to the Department of Electrical Engineering and Computer Science
on August 5, 1988 in partial fulfillment of the requirements for the Degree of
Master of Science in Technology and Policy.

ABSTRACT

The principles of design are universal to all disciplines and apply in all contexts. This thesis is maintained and advanced in three parts.

First, two design principles or axioms are introduced and placed in the context of a broader philosophy of design which subsequently evolves into a description of a design science. The purpose of this first part is to show that the design axioms presented here are rooted in philosophical principles and are an important part of a complete design science.

Second, the validity of the axioms are tested through their application to the design of a filter circuit. The design methods of Taguchi are utilized to optimize this circuit and the principles underlying the application of these methods are shown to be consistent with the axioms of design. Thus, the axioms are consistent with proven methods of engineering design.

Third, the validity of the axioms are tested by applying them to the design of the United States Patent System. A system design representation is developed for the patent system from which the application of the axioms is facilitated. The application of the axioms yielded findings and recommendations for some intellectual property policy questions of immediate concern in 1988. The successful application of these design axioms to a system of laws and policies illustrates that they are context independent.

Their are three main audiences for this thesis. First, educators in the field of design, whether it be engineering design, organizational design or policy design will benefit from understanding the universality of design and its interdisciplinary nature. Second, design engineers interested in the application of design principles using the statistical techniques of Taguchi will find the second chapter most useful. Third, those involved in the design of the United States patent system may find the systems perspective presented in chapter III to be useful in making policy decisions and laws.

Thesis Supervisor: Nam P. Suh, Professor of Mechanical Engineering

ACKNOWLEDGEMENTS

The inspiration for this thesis began more than five years ago, when as a young engineer I began to seriously contemplate the meaning of my profession and its relationship to society. I was referred by a friend to speak with Professor Larry Bucciarelli at MIT concerning this topic, at which time I was introduced to the MIT Technology and Policy Program. I would like to especially acknowledge Professor Bucciarelli for this introduction to MIT as well as his continued support and contribution to this thesis as a reader and constructive critic.

Professors willing to step outside the traditional bounds of their disciplines in support of interdisciplinary theses are not the most abundant resource at MIT. For his willingness and desire to expand and not constrict engineering disciplines, I am indebted to my thesis advisor, Professor Nam Suh. In addition, the design axioms developed by Professor Suh and his colleagues contributed in large part to the argument developed in this thesis. For this contribution, as well as his continued encouragement, I would like to acknowledge and thank Professor Suh.

For his contribution as a reader of this thesis, and particularly for his significant help in my understanding of the Taguchi methods presented in Chapter II, I would like to thank and acknowledge Professor Don Clausing.

In addition to the academic support and encouragement necessary for the completion of this thesis, I would like to acknowledge the financial support and moral encouragement I received from family, friends and colleagues during the past two years.

Professor Charlie Sodini gave me my first opportunity at MIT as a teaching assistant, without which I could not have supported my education, in which I learned a great deal and for which I am sincerely grateful.

Professor Richard Lester gave me the opportunity to work as a research assistant on one of MIT's most significant and unique interdisciplinary studies ever. Without his support and encouragement to pursue my vision, this thesis would not have been possible.

I would also like to acknowledge Dean Alfred Keil for the early encouragement and comments he provided during my initial thesis formulation. Thank you Dean Keil.

The financial sacrifice of my two years at MIT was felt in some way by each member of my family. I would like to especially acknowledge my mother Joan, my grandmother Sarah, my two brothers Robert and Gerald, and my two sisters Maryanne and Susan for their sacrifice on my behalf and for the solace gained by their presence. Also, for his contribution as a mentor and friend, I would like to thank the Reverend A. Paul Gallivan.

Deo Gratias.

Stephen F. Filippone

PREFACE

Toward A Better Understanding Of Design Methods And Their Elemental Principles

Engineering is a method for affecting change, which from a historical perspective is perhaps the oldest profession. From before the beginning of written history people have attempted to improve their lot in life through methods which are commonly referred to today as engineering. Prehistoric people developed tools from stone for hunting, the Egyptians first forged copper for weapons and the Romans built aqueducts to supply water to their vast cities. The ancient Greeks, the Romans and the Egyptians all designed great civilizations and systems of governments. Today the products of engineering(e.g., computers, airplanes, policies, etc.) affect our daily lives through law, economics, medicine, politics, religion and science. These engineering feats and others throughout history have changed the lives of the masses, whether for better or worse is left for the historians to decide. However, recognition of this history of change through engineering methods is an important first step in comprehending the nature of the profession.

The engineering method is a strategy for change which can and should be, an activity in which people of all backgrounds participate. Koen(1985)* describes the engineering method as "the strategy for causing the best change in a poorly understood or uncertain situation within the available resources; by reason, I mean the 'ability to distinguish between the true and the false' or what Descartes has called 'good sense'." People today tend to focus too often on the objects of technology (e.g., computers, bridges, robots, etc.) and associate these objects with engineering, when the real distinction is in the method through which these objects are developed. It is understanding the method and the principles behind them which distinguishes the engineer from the non-engineer. It is not one's grasp of science, mathematics, technology, philosophy or any other specific discipline alone that makes good engineering, but rather the method through which one uses this knowledge in solving the problems of human existence. It is toward a better understanding of these engineering design methods and their elemental principles which is the primary motivation for this thesis.

The reason for stressing the importance of the engineering design method is gained from an understanding of what engineers do and how they interact with nature and society in the course of their work. There are many perceptions of what engineers do which suggest the importance of this technical-societal relationship, however few have attempted to delve more deeply to explain a rigorous approach to engineering design that would give the engineer a tool for dealing with this often nebulous relationship. To better understand how a rigorous, structured approach to design could incorporate this unstructured, technical-societal relationship within its format is one goal of this thesis.

From the multitudes of perceptions we encounter of what engineering is, the question remains for the student of engineering: what is engineering and how is it practiced? Billington(1986)** says that the "central activity of engineering is design, and the primary motive for design is the creation of an object that works." Furthermore in defining engineering, one must distinguish it from science, as the two are different. "Science is discovery, engineering is design. Scientists study the natural, engineers create the artificial. Scientists create general theories out of observed data; engineers make things, often using only very approximate theories."

Engineering design requires the ability to understand or come to know problems that manifest themselves in an unstructured as well as a structured format. Designers must have the ability to think in an uncertain, risky and increasingly complex world. They must formulate solutions to problems that are not yet completely understood, requiring a certain adaptability in order to deal consciously and responsibly with the unknown.

Engineers do not have the luxury of the scientist to merely seek the truth, but must present solutions to problems in which the complete set of facts or truth is not known. Thus engineering design is a learning process as much as it is a method to produce problem solutions.

During the initial phase of the engineering design process, the known facts concerning a given problem are sought. At this time the engineer must compile information from unstructured and sometimes contradicting sources(e.g., the environment, consumer reports, colleagues, journals, etc...) and form the best understanding of the problem possible, given the information constraints inherent in a complex and dynamic environment.

The designer then forms a conceptual understanding of the problem from which possible solutions are envisioned.

Imbedded in these alternative solutions are constraints based on the designers' understanding of the problem, in which values are inherently present. Within these constraints the designer can now comfortably shift back to a more structured analytical approach to evaluate the alternatives. There are numerous techniques to evaluate all sorts of engineering solutions, from systems analysis to circuit optimization techniques. The engineer chooses an appropriate technique for evaluation and determines a dominant solution among the alternatives. The engineer may choose the most familiar technique or the one valued as being the best. In either case a value judgement is made at this phase of the design which will often influence the choice of dominant design. This dominant solution is based on information obtained earlier in the initial stages of the design process and its validity is thus based on two assumptions.

First, that the frame of information from which the possible solutions were derived is complete enough in depth and breadth so that the analysis brings out the majority of pitfalls inherent in any alternative, especially the most crucial ones. The enormous amount of information available today and the resources available for its collection make the gathering of all the relevant information an easier task now than in the past. However, engineers can never be sure about their products until they have been fully tested and thus the weakness in design brought on by this first assumption is something engineers throughout history have learned to deal with by testing.

Second, the dominant solution derived from the previously obtained information assumes that the world from which this initial frame of information was extracted has not changed considerably in a manner that would appreciably affect the solution. However, engineering solutions take time to analyze and develop. During this time, the world of information may change, making it necessary to reevaluate the solutions and redefine the problem. In order to develop solutions with lasting value, the designer must shift between the structured analysis phase and the unstructured problem definition and solution conception phases of design.

The need for this ability has been brought on by the fast rate of technological as well as cultural change in our society today. Our culture has become increasingly technology driven. With communication links possible to all parts of the world, people are evolving culturally faster than ever before. They are learning new desires for change and engineering is the primary tool in their quest. Because the desires of some may inhibit the desires of others, engineers have developed technologies that satisfy some and not others(e.g., nuclear power vs. the environment, computer power vs. privacy, etc.).

This conflict in desires results from a complex world in which large variations in values are common in many segments of society. Design engineers in all disciplines(e.g., electrical, mechanical, political) need to incorporate a way of understanding these conflicts within their design methods if they are to provide a means toward the resolution of such conflicts. One goal of this thesis is to provide some direction for engineers in the quest of such ends.

Currently design is thought by many as not worthy of being a discipline itself, but rather is taught within the structure of other disciplines. Mechanical engineering teaches mechanical design courses, electrical engineering teaches circuit design courses, management teaches organization design courses, and political science teaches policy design courses. This thesis should contribute to the better understanding of engineering design as an activity that transcends the traditional disciplinary bounds.

More practically this thesis should place design principles and methods within the framework of the science of design to give practitioners a base from which to view their work. In addition, the better understanding of a generic design process will give practitioners of different disciplines a basis for more fruitful discourse. In this light, the advancement of the science of design has great potential for increased interdisciplinary activity. The audience for this thesis are all those who seek to design, but more importantly, those who strive to teach design.

---

* Koen, Billy V., Definition of the Engineering Method, American Society for Engineering Education, Washington D.C., 1985, p.5.

** Billington, David P., "In Defence of Engineers", The Bridge, Summer 1986, p.4.

TABLE OF CONTENTS

CHAPTER I - A PHILOSOPHY OF DESIGN

For to be possessed of good mental powers is not sufficient;
the principle matter is to apply them well.

René Descartes, Discourse on Method

1.1 Introduction

What is design? Why design? It is through struggling with these questions in the search for their meaning and truth that, will in the end, result in designed artifacts or objects for the benefit of society. The purpose of design is to **create an object** that **works** based on a set of **defined requirements**. In the context of engineering design, an additional purpose is to define the requirements in such a way that the designed object provides some **utility for society**. The understanding of what it means to design can be broken into those three basic components; (1) to create an object, (2) to make that object work as defined, and (3) to provide utility to others. The process of design requires the constant reflection on all of these components.

To create a physical object is not meant in the genesis sense of forming the existence of something that was not, but rather in the reforming of physical objects that already exists. To exist in the Descartes sense is to think and through thinking we come to know, and as thought produces knowledge, knowledge creates objects. Physical objects are not created, but rather creativity takes place during the thought process where new relationships between physical objects are uncovered. Thus, in understanding this component of what it means to design, the theory of knowledge and the thought process of the mind need to be understood at a less abstract level than has been put forth by Descartes, Bacon and Leibniz, as well as the more contemporary philosophers. A more practical theory (or theories) of the representation of knowledge for complex design problems is needed to aid the designer in the creation of an object that works and that has utility.

To make an object work, one must obviously be able to distinguish between something that works and something that doesn't work. To proceed in this endeavor the designer must be armed with a set of functional goals that are consistent throughout in purpose. By formalizing knowledge in a logical and/or sometimes random process, these functional goals become working objects. Logical formalization techniques are often called methods while the random formalization of knowledge is often called creativity. Whichever formalization process is chosen, the end result must be compared with the particular functional goals to determine if the design works.

To provide utility to society or the specific customer can be thought of as the most fundamental goal of any design that will, in the final evaluation, distinguish an object that works from one that doesn't. There are degrees of utility which makes the distinction between objects sometimes vague and one must choose between two objects that work in which the one offering the greater utility is not obvious. Constraints can sometimes force one to choose between the better of two non-ideal designs. How one makes such distinctions has been the subject of healthy inquiry on the part of academics, politicians, engineers and all professions that design(i.e., all professions). More basic however, is the recognition that such distinctions can not be made by the designer, but rather by those for whom the object is designed(i.e., the customer) as well as by those in the general public that will be affected by such an object(i.e., society). Thus construed, the distinguishing of objects that provide utility to others is a social choice rather than a design choice.

These three components of what it means to design objects (i.e., to create, to work, to provide utility) are proposed here as the framework on which a more detailed philosophy of design rests and from which the science of design has emerged. The term object has been used here to describe that which is the subject of the design and it may or may not be a material object; it could be a law, policy, electric circuit, sales plan, cutting machine, etc. Thus in a philosophy or science of design, the universality of its application to all domains of knowledge is essential.

Good design practice requires the constant reflection on all of these three components of design philosophy. Failure in design prevails when any of these components are ignored as is the case of a product design in which the voice of the customer is ignored. In this case, when the question of utility to the customer is decided by the designer in determining the functional goals of the product, even when the goals are achieved in an object that works, the product will not work in the eyes of the customer and will thus be judged a failure. In this situation, when the designer fails to reflect the utility of the customer in his product, poor designs expound. Competition in design today rests in large part on the ability of the designer to satisfy the needs of the customer.

Design failures also occur when an object is created for which a need is evident and reflected deeply in the functional goals of a design, but for which the physical embodiment of the created object does not work. This problem is common in designing laws and policies as one can sometimes easily identify needs and formalize general statements into a law or policy. However, when the context of the implementation for such laws and policies is not reflected upon in their design, failure abounds.

More detailed examples of such design failures will be examined in the following chapters through the application of two fundamental principles of design (or design axioms) to the United States Patent System. The two axioms presented in this thesis were developed over the past eleven years by researchers at the Massachusetts Institute of Technology for the purpose of developing a disciplinary base for design and to teach students generalizable knowledge that is fundamental to good design practice. Under the direction of Suh (1988) these two basic principles have been developed and in their current declarative form can be stated as:

Axiom 1. The Independence Axiom
Maintain the independence of functional requirements.

Axiom 2. The Information Axiom
Minimize the Information Content.

The first part of this thesis will be devoted to placing these axioms in the broader context of a design philosophy or science. The purpose of this is to illustrate the validity and utility of axiomatics in the context of a design science while at the same time illustrating that these axioms are not the whole of such a science.

The distinction between a design science and a philosophy of design is that the philosophy describes what is known and how, while the purpose of the science is to build upon the knowledge base of the philosophy. This distinction is very vague and will not be expounded upon further here, as such an argument would have no purpose in this thesis. Thus, the terms philosophy and science will be used interchangeably for the remainder of this opus.

In the second part of this thesis a contribution to the science of design will be made through the comparison of design axioms and methods. The distinction between axioms and methods will be made in the first part of this thesis as they fit within the context of a design science. In this second part the distinction will be elaborated upon with a more detailed comparison of the two axioms stated above and the Taguchi Methods of engineering design which were originated by Taguchi(1987) of Japan. The purpose of this comparison is to test the validity of the axioms against a proven method in engineering design. The truth of the axioms rests not on a proof but rather on the lack of any evidence that would refute them. Thus, in this second part, evidence based on the practical application of Taguchi methods will be probed for substantiating as well as refuting evidence.

In the third part of this thesis, the axioms will be applied to the design of the United States Patent System. The purpose of this application is to test these axioms in a context outside the realm of engineering. Nam Suh (1988) provides many examples of engineering applications of these principles as well as an organization design application. However, this thesis represents the first formal application of these axioms to a system of policies and laws. The Patent System itself is devoted to the initiation of innovation and designs that provide utility to others, giving this application of design axioms an interesting twist since they are devoted to the same initiatives. It is also the goal of this section to provide useful recommendations to policy and law makers in the area of intellectual property rights.

This work consists, therefore, of three parts each with its own purpose and each contributing to the overall thesis, that within the context of a general science of design there exists a set of basic principles universal to the application of good design practice. Thus construed, the science of design represents a discipline of study of which its utility is common among all professions.

1.2 A Philosophy of Design

In defining a design philosophy it is not important, or perhaps even possible to precisely state a definition beyond reproach. This however, should not inhibit one from striving to understand and better comprehend the philosophy of design in the pursuit of an improved design practice. It is this pursuit that guides this thesis from the description of a broad philosophy of design to the practical implementation of its formalize principles.

A design philosophy is meant to provide insight into an approach or a starting point from which questions can be posed. A primary goal for such a philosophy is to aid in the posing of the right set of questions. Design failures prevail when an approach becomes so narrow in scope that the designer fails to ask the proper questions.

The 1979 crash of the DC-10 on take-off from Chicago's O'Hare Airport can be cited as a design failure that resulted from the posing of an incomplete set of design questions. The engines in the DC-10's are connected to the wing through a pin assembly, that when removed for servicing and replaced in a manner unanticipated by the designer, caused the flange into which the link fit to tear around its pin-hole. After the cracked flange was identified as the cause of the fatal crash, several other DC-10's were found to have the same failure that were subsequently repaired. The designer failed to ask the necessary questions and define functional requirements that allowed for the excessive abuse received by the flange during maintenance.

Petroski (1985) explains that such failures offer an invaluable lesson to the designer which play an important role in design. Although it is doubtlessly true that to learn from ones mistakes is an essential virtue for designers, equally or perhaps more virtuous is the ability to obviate such mistakes through the posing of a complete set of initial questions. Thus, failures are often not the result of the inability to solve a problem but rather the inability to recognize problems.

In the course of logical inquiry, designers must be able to step back and forth from narrowly defined goals such as designing the most powerful engine, or the fastest microprocessor to a broader perspective that asks fundamental questions of purpose, safety and overall utility to humankind, which are necessary for good design practice. The need to function in both a detailed, structured, and analytical world, as well as an often broad, unstructured, and loosely defined world is an important aspect of design practice which becomes evident with an understanding of the philosophy or science of design.

George (1981) describes philosophy as having three distinct banners:

1. That knowledge must be uncertain and contextually dependent.
2. That formalization, say by definition, diminishes the uncertainty and contextual-dependence.
3. That we can, and in the end should, normally think of all philosophical analysis as occurring within a behavioral context.

These banners will serve as the premise for a framework from which the philosophy of design will be explicated and the elements of a design science can be placed. The following premises are posed:

1. There is uncertainty and interdependence in all information.
2. Formalization of design, say by axiomatizing, diminishes uncertainty and disciplinary-dependence, but never completely.
3. Design should be thought of as occurring in a social context.

The first premise is concerned with the representation of knowledge (or information), the second with formalization of such knowledge and the third with the recognition of design occurring within a social context.

1.2.1 Representation of Knowledge

Epistemology is that part of philosophy concerned with what we know and how we know it. Clarke (1982) describes knowledge as "those truths which must be known as a condition for the possibility of knowing anything else". Certainly it can be said that knowledge comes from a growth of experiences through which one comes to know or understand in a variety of ways and contexts.

Snyder (1970) in his analysis of how students come to know or understand, shows that it is dependent upon the context in which information is presented. He explains that by holding to a single model of the world while ignoring the changing context of the world environment, the opportunity to know, to grow, and to change will be forfeited. Thus, any representation of knowledge in the context of design must have imbedded in its structure a flexibility to allow it to change itself in structure as well as in content.

A designer must be flexible in the understanding that, what is known changes with time as circumstances change. Even Newton's Laws of Motion published in his Principia of 1687 were known and accepted for more than 200 years before Einstein in 1905 explained that they were invalid for objects travelling with velocities close to the speed of light. The failure of our knowledge about objects in motion to explain electromagnetic waves travelling at velocities close to the speed of light is clear evidence that knowledge is context dependent, no matter how well established. Another problem of particular importance that is faced by the designer is the representation of uncertainty in the available knowledge or information. Information is in general uncertain and the extent to which this uncertainty can be understood will guide the designer in choosing between alternatives.

Probability techniques are often used to measure the certainty of information. For example, an engineer may estimate the probability of a circuit failure based on the knowledge of the tolerances for the specified parts. An insurance company estimates the probability of accidents occurring based on histories of past accidents and designs policy plans based on these estimates. All such estimates are uncertain and thus designs based on them are subject to some level of risk that they will perhaps fail.

In the light of the first premise that information is uncertain and interdependent, figure 1.1 is a spatial representation for information in design that will be utilized later in applying the designs axioms. As one proceeds to move between the unknown, unstructured, functional, physical, and process spaces, information is formalized to produce the object of the design.

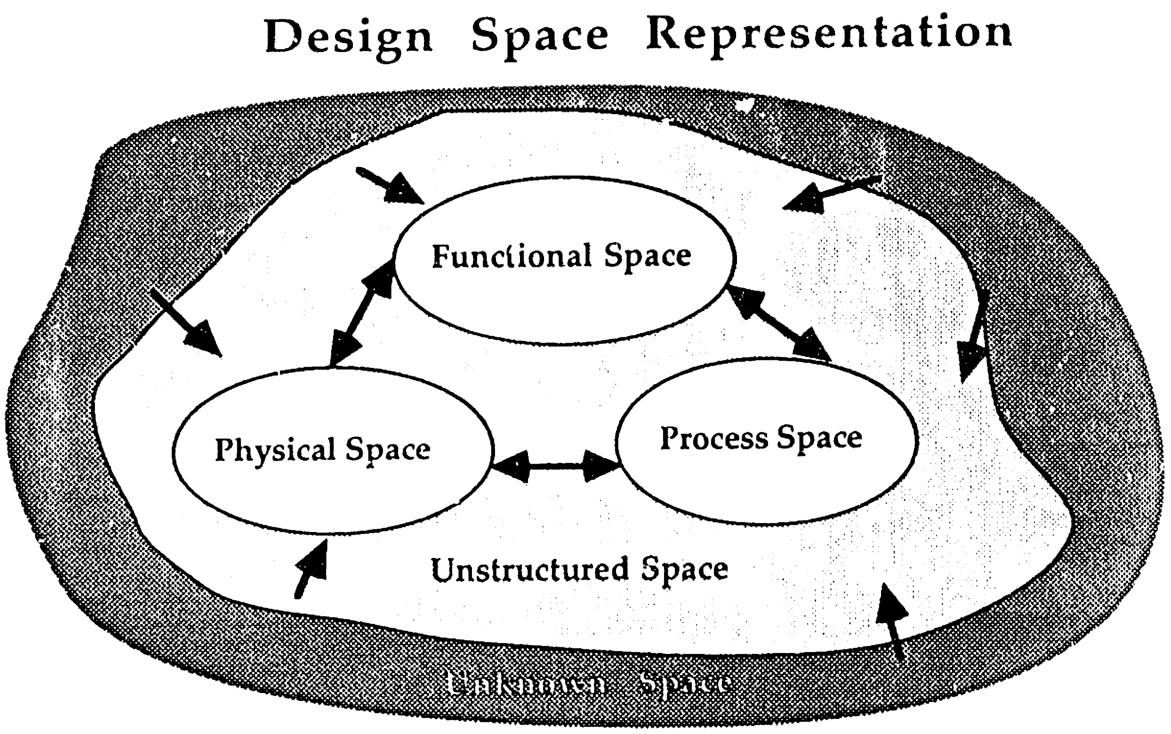


Figure 1.1 A spatial representation of the design process illustrating the causal interaction between the functional, physical, process, unstructured and the unknown domains of information.

The unknown space is depicted as surrounding the four known spaces and represents knowledge that has not yet been discovered. As experience generates new knowledge of the world, the unstructured space grows.

The unstructured space represents the domain of all knowledge that is available for the designer to formalize in the context of a particular design. The information included in this domain is vast and growing, it includes knowledge of the physical sciences, social science, customer needs, societal values, law, politics, and any other realm of understanding that would potentially impact the success of a design. Depending on the particular object of design, some of these realms of understanding will be more important than others. Nevertheless, the designer is often faced with a large unstructured mass of knowledge that must be formalized into the description of an object that works.

The process of formalizing knowledge into the object of design is depicted as taking place within the functional, physical, and process spaces. During this process, the lack of information within the unstructured space may lead to the necessary inquiry into the unknown. Such an inquiry may be done scientifically and methodically or randomly and creatively.

The functional space defines the basic requirements for the object of design. It is in moving from the unstructured to the functional space that the conception of a design takes place. The determination of functional requirements directs the inquiry into the unknown and the analysis of the unstructured as one proceeds to develop the physical embodiment of the construed functional description.

For example, if in the conception of functional requirements for a power generating plant, the knowledge concerning the formation of acid rain in the atmosphere, the hazards of nuclear waste and the dependence of the US on foreign oil supplies were all considered conditions to be avoided, inquiries would be directed toward alleviating such conditions. If the proper set of functional requirements are not considered, design failures prevail as in this case of the power generating plant one might create acid contaminated waterfall, radioactive contamination of our ground water or an overwhelming economic burden on our nation. Thus, in the forming of functional requirements, the context in which the object of design is to be used and all knowledge pertaining thereto must be considered.

The physical space represents the formation of knowledge needed to achieve the stated requirements in the functional space. For engineering artifacts, parameters such as dimensions, temperature, voltage, tension, and any other physical parameter that may describe the embodiment of a design are defined in this space. In the patent system, the physical parameters would be the statements of law and policy that define the system. Thus, in the physical space, the embodiment of the design is described.

The process space represents the formation of knowledge needed to manufacture or implement the embodiment of a design in the physical space. Design failure abounds when the process parameters needed to manufacture a given embodiment described by its physical parameters are unknown or uncertain. Information in the process space is represented typically as known parameters of the machine, individual or group of individuals that are responsible for the manufacture or implementation of the design.

The design process thus construed consists of a continuous movement between the design spaces as one proceeds from the most general functional goal of a design to its most detailed description and manufacture. This organization of design information can be represented hierarchically in which the more general functional requirements decided early in the design become constraints on all subsequent design parameters. It is through the formalization of knowledge in the functional, physical, and process spaces that a design is conceived, described, and produced.

1.2.2 Formalization of Knowledge

As explained in the previous discussions, knowledge or information is uncertain and context-dependent. The goal in formalizing what is known and how it is known is to reduce the uncertainty and context-dependency of the information. How one approaches the formalization of information is an important aspect of a design philosophy which will be introduced here with a discussion of two ways to formalize design information; axioms and methods.

In formalizing knowledge, one organizes it in a manner that directs our further inquiry of additional information in a logical process. The additional information serves to reduce the uncertainty of the design. Without an organized information search, one could be looking for a needle in a hay stack when faced with the explosion of information in todays world.

When a solution or idea is conceived through a process of random inquiry, it is called creative. An individuals thought process or path of inquiry can never be totally directed, and in fact should not, but to the extent that it is, creativity can be inhibited when that direction is too narrowly focused. The design thought process should be goal directed rather than disciplinary directed, which stimulates rather than inhibits creativity.

When knowledge is abstracted from the world about design, there exists the danger that analysis will be carried out such that the details of the context in which the design information will be used is forgotten. When this occurs, the logical consequences of the analysis for the assumed context may be false and the design may fail. These dangers of formalizing design information into an object that fulfills a need must be avoided.

Axioms are formal statements or postulates derived from what is known. Since uncertainty abounds in all information, an axiom can never be proven as absolute truth, but nor can it be refuted. Axioms are thought to represent self-evident truths and since they can not be logically proven, the interest is predominantly in their logical consequences or theorems.

Axioms are purposeful statements and thus form knowledge with a specific goal in mind. In the case of Euclid(300 B.C.) the application of the knowledge of geometry was the goal and in the case of Suh(1988) the application of the knowledge governing design is the goal. Before knowledge can be formalized into axioms, it must be known and understood at some sufficient level. Because we have not reached that level of knowledge for all purposes, axioms have yet to be postulated in many spheres of knowledge. For example, we lack axioms for many purposes such as axioms of psychology, axioms of politics, or axioms of management. It is conceivable that axioms do exist for such knowledge areas, however sufficient and consistent information has not yet been observed or organized in a manner that would elucidate such axioms.

Once knowledge has been formalized into axioms, they should be tested for completeness, consistency, and appropriateness to the intended application. Subsequent chapters in this thesis are devoted to testing the axioms of design. Chapter II will describe axioms of design in more detail and comment on their relation to the philosophy of design. Subsequent chapters will illustrate the utility of these axioms through their application.

Methods. Within the realm of design there are, as in any philosophy, numerous methods posed from which to deduce results. These methods must be viewed from the context of their intended purpose. One aim of the philosophy of design would be to distinguish between correct and incorrect methods. There is room for many correct methods within a design philosophy and, as part of a design science the relationship between and implications of such methods should be explored, realizing that our results are in part an artifact of the method employed in their attainment.

A.D. Woozley (1949) expresses the matter concerning the controversy surrounding different philosophies, which can be adopted here to express the controversy between various design methods as a healthy activity. That is, controversy is desirable; only unreflective prejudice and passive acquiescence are to be avoided. Thus, the healthy controversy or discussion concerning the differences between various design methods is one important aspect in the science of design.

This thesis will contribute to the science of design in part by examining the relationship between the proposition of design axioms and Taguchi methods. In choosing this particular method it should not be inferred that the author advocates unequivocally the use of such methods in all situations. However, a better understanding of the relationships between any methods (proven in some context) will contribute to our understanding of design science and consequently to a better philosophy of design. I also believe the reverse to be true, that a better design philosophy leads to a more complete design science.

Rule IV of Descartes' "Rules for the Direction of the Mind", which has been preserved by Eaton (1927, pp. 48-49), states the need of a method for finding out the truth. Descartes says that,

> ...by a method I mean certain and simple rules, such that, if a man observes them accurately, he shall never assume what is false as true, and will never spend his mental efforts to no purpose, but will always gradually increase his knowledge and so arrive at a true understanding of all that does not surpass his powers.

Descartes is speaking of a general method to guide ones thought process in the acquisition of purposeful knowledge, which is what 'design methods' also aspires to do. For example, Beakley and Chilton (1973) describe a method to guide a designer that entails the acquisition of knowledge in a logical progression through three phases; (1) the feasibility study, (2) the preliminary design, and (3) the detail design. The context of their method is the engineering profession for which they describe its purpose "is to develop technical devices, services, and systems, for the use and benefit of man".

Thus methods are purposeful, for Descartes it was the acquisition of the knowledge of all things for which the mind was capable, while for Beakley and Chilton it was somewhat more narrowly focused on the development of technical devices. Methods that rely on context-dependent knowledge are to that same degree context-dependent. Consequently, methods developed for designing integrated circuits are not necessarily useful in designing large centrifugal pumps. Similarly, methods developed for designing engineering artifacts may not apply to designing policies and laws. What is sought in this thesis is a set of general principles that would apply in all contexts of design.

Descartes' "certain and simple rules" are not meant as a set of general principles that can be learned and mechanically applied to yield truth, as can be said of axioms. This is the key distinction between a method and an axiom, that is, a method guides in the acquisition of truth while an axiom is a true statement.

The formalization of knowledge into methods that are context-dependent or into axioms that are statements of truth are major subjects within a philosophy of design. Such formalizations serve to reduce uncertainty in design by guiding the acquisition of purposeful knowledge. The study of the relationship between various formalizations of knowledge (e.g. methods, axioms, rules) will ultimately reveal more truths in a philosophy of design and consequently better formalizations. The study of these relationships is the task of a design science.

1.2.3 Societal Context of Design

All design is the product of human labor (at least until computers are given imaginations), and the context within which people design varies between countries, cultures, organizations, and individuals. The designed artifact will depend not only on what is known (i.e. representation of knowledge) and how knowledge is applied (i.e. formalization of knowledge), but also on the perspective from which knowledge is viewed. There are many factors influencing such a perspective that vary from cultural backgrounds to interpersonal skills.

Because of the multiplicity of such factors and the variation between individuals, this aspect of a design philosophy perhaps presents the greatest challenge to formalization. The unstructured variation in individual cultures and interpersonal skills make it difficult to study their influence on the whats, hows, and whys of the designs process. However, the understanding of these influences on design is necessary for a complete philosophy of design.

Hubka (1985, p. 134) lists seven factors influencing the design process of which three would be considered within the behavioral-context. These factors include the quality of design engineer, the management of the design process, and working conditions. The quality of the designer depends on education and the ability to convey information concerning design science. The management of the design process requires the coordination of many individuals and activities, and providing the proper working conditions requires the knowledge of human needs and motivations.

Hubka's other four factors include; working means, technical information, design representations, and design methods. These factors are consistent with the philosophy of design presented here and are contained within what has been previously described as the representation of knowledge and the formalization of knowledge.

Bucciarelli (1988) has studied the design process within two organizations as a participant observer and concludes that design is a social process. This social process includes the activity of individuals engaged in design both within and external to the confines of the firm. It includes the study of how individual views, beliefs, interests, and disciplinary competencies contend, and are resolved and reflected in the process of design.

1.3 The Need for a Design Science

Hubka (1985, p. 136) describes design science making no distinction between philosophy or theory as,

> Design Science (or Science of design, or design theory, or design philosophy) can be defined as a collection (system) of logically connected knowledge of engineering design. Consequently design science is the most complete area of engineering design covering the whole field of problems.

The science of design is unique among scientific disciplines in that it seeks knowledge which is common throughout the whole of all science. Descartes, in "The Rules for the Direction of the Mind", which has been preserved by Eaton (1927, p. 51), touches on the goal of the science of design while expounding upon his "method for finding the truth".

> Such a science should contain the primary rudiments of human reason, and its province ought to extend to the eliciting of true results in every subject.

The science of design which here has evolved as a natural progression from a philosophy must aim toward, as Descartes exclaims, results in all domains of knowledge. This concept of universal applicability and utility of a design science is still a strange and vague concept for many, three centuries after Descartes. For others however, such as Simon(1981), Hubka(1985) and Suh(1988) there does exist such a science, the elements of which have been outlined in Table 1.1 as emanating from the three basic elements of design philosophy: (1) the representation of design knowledge, (2) the formalization of knowledge, and (3) the recognition of the social context of design.

Table 1.1 - The Science of Design

Representation of Knowledge	Formalization of Knowledge	Social Context
1. Theory of structure and design organization 2. Representation of design problems. 3. Creativity Theory 4. . . . 5. . . .	1. Theory of evaluation -Utility Theory -Statistical Decision Theory -Information Theory 2. Formal Logic of Design -Axioms, Rules, Theorems -Methodologies 3. Study of the relationship between methods of formalizing knowledge. 4. Computational Methods -Linear Programming -Dynamic Programming -Control Theory	1. Organizational Behavior -Theory of Group Dynamics -Interpersonal Relations 2. Management Theory 3. Human Factors 4. Ethics 5. . . .

1.4 References

Bucciarelli, Louis I., "An ethnographic perspective on engineering design", 1988.

Clarke, Desmond M., Descartes' philosophy of science, Pennsylvania State University Press, University Park, 1982, p.81.

Descartes, René, Descartes Selections, Ed. Eaton, Ralph M., Charles Scribner's Sons, New York, 1955.

George, F.H., The Science of Philosophy, Gordon and Breach, Science Publishers, Inc., 1981, p.22.

Hubka, Vladimir, "Attempts and Possibilities For Rationalization of Engineering Design", Design and Synthesis, Elsevier Science Publishers B.V., North-Holland, 1985.

Petroski, Henry, To Engineer is Human, The Role of Failure in Successful Design, St. Martin's Press, NY, 1985.

Simon, Herbert A., The Sciences of the Artificial, MIT Press, Cambridge, 1981.

Snyder, Benson R., The Hidden Curriculum, MIT Press, Cambridge, 1973.

Suh, Nam P., The Principles of Design, Oxford Press, 1988.

Taguchi, Genichi, System of Experimental Design, Engineering Methods to Optimize Quality and Minimize Cost, American Supplier Institute, Inc., 1987.

Woozley, A.D., Theory of Knowledge, Hutchinson, London, 1949.

CHAPTER II - THE DESIGN AXIOMS USED IN CONJUNCTION WITH THE TAGUCHI METHODS OF ENGINEERING DESIGN

2.1 Introduction

The purpose of this chapter is to compare the methods of Taguchi(1987), in the context of engineering design, with the two axioms of design uncovered by Suh(1988). If these axioms are true, they should not contradict other methods in principle. To proceed in this analysis, first the axioms and their implicit assumptions will be described in detail. Second, an example of a simple passive filter design will be worked through using the analysis of variance and signal-to-noise metrics described by Taguchi(1987). This example will serve as a basis for comparison from which the basic principles underlying the Taguchi methods can be uncovered and compared to the design axioms.

2.2 Axioms of Design

The design axioms described here can be defined as the set of fundamental principles that guide a designer in the formation of an object that works, as specified by a given set of functional requirements. In fact, the functional requirements which have been previously defined serve as a basis for which to judge if the final object works or not. Thus, the axioms are not meant to guide the designer in choosing the correct set of functional requirements, but rather in the formation of an object that can fulfill the given requirements.

However, if the physical object needed to fulfill the given functional requirements can not be found, then logical inquiry into alternate functional requirements is often spurred. In this way the axioms can serve to decide when new or different functional requirements may be necessary. Thus the axioms as presented here are put in the perspective of governing the movement between the functional, physical and process spaces of figure 1.1. The business of defining functional requirements will be taken up with the Patent System design in chapter III, but for the purposes of this comparison, the example illustrated here will begin with a given set of functional requirements.

Although the formalization of knowledge in the unstructured and unknown spaces are certainly part of the design process, the fundamental principles governing the application of such information have not yet been uncovered. Thus, the two axioms presented here may only be a subset of the total set of design axioms. In Chapter I the design axioms were placed within the context of the entire design process so that their meaning and implicit assumptions can be described in detail while keeping them in proper perspective.

2.2.1 Independence Axiom

The first axiom is the independence axiom which simply states:

> Maintain the independence of functional requirements.

This statement means that as the physical embodiment for the object of design is specified in the physical space and manufactured or implemented in the process space, each functional requirement should be independently controlled by a set of physical and process parameters. For example, the set of physical and process parameters designed to fulfill functional requirement 1 (FR_1) should not affect functional requirement 2 (FR_2) and the parameters designed to fulfill FR_2 should not effect FR_1. When this occurs, the functional requirements are said to be uncoupled.

An uncoupled design in this sense does not imply that one needs a different physical object to fulfill each functional requirement, but rather a different physical parameter. Physical objects often have many physical parameters that can potentially fulfill various functional requirements. For example, two functional requirements for the body of an automobile may be to avoid rust and to be aesthetically pleasing. Both of these requirements can be fulfilled with one physical object (e.g., paint) which has two physical parameters (e.g., color and impermeability to water). In this sense the two functional requirements are uncoupled in the physical world.

Autobody Design

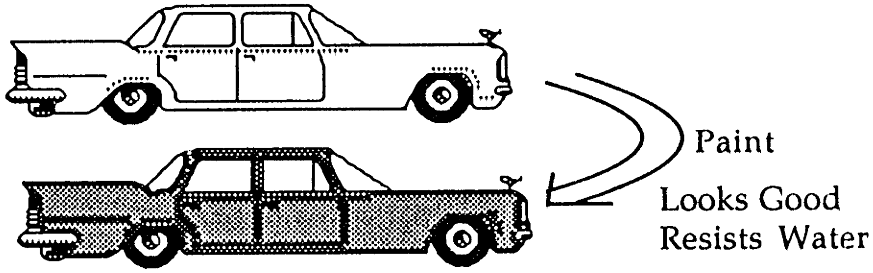


Figure 2.1 The functional requirements to look good and to resist water for the body of an automobile can be achieved with the same physical object (e.g., paint).

The concept of functional independence can be further clarified through a the use of a mathematical representation for a design as shown below in figure 2.2.

Design Matrix Representation

$$\begin{array}{ccccc} \textbf{Functional} & & \text{Design Matrix} & & \textbf{Physical} \\ \begin{Bmatrix} FR_1 \\ FR_2 \\ \cdot \\ \cdot \\ \cdot \\ FR_n \end{Bmatrix} & = & \begin{bmatrix} A_{11} & A_{12} \cdots A_{1n} \\ A_{21} & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ A_{n1} \cdots\cdots & A_{nn} \end{bmatrix} & \text{X} & \begin{Bmatrix} PP_1 \\ PP_2 \\ \cdot \\ \cdot \\ \cdot \\ PP_n \end{Bmatrix} \end{array}$$

Figure 2.2 A mathematical representation for a design showing the mapping of the functional requirements into the physical parameters via the design matrix A.

In figure 2.2, FR_1 can be said to be totally independent of all other FRs when all design matrix elements in the first row, with the exception of A_{11}, are zero. Similarly, FR_n can be said to be independent of all other FRs when all design matrix elements in the n^{th} row, with the exception of A_{nn}, are zero. When all functional requirements are independent of each other, the design is said to be uncoupled and the resultant matrix is a diagonal.

A metric for the representation of the matrix elements Aij is suggested by Suh(1988) as the change in FRi with respect to PPj, which is then normalized by dividing by FRi. This can be formally stated as follows,

$$A_{ij} = \delta FR_i/\delta PP_j)/FR_i \qquad [2.1]$$

Each functional requirement is a function of the physical parameters, as well as unpredictable noise parameters. In design it is important to be able to accurately predict the outcome of a FR, which makes it necessary to know how a FR changes with respect to changes in physical parameters and noise. The change in FR_i can be represented as,

$$\Delta FRi = \delta FRi/\delta PPi + (\Sigma_j \ \delta FRi/\delta PPj)_{j \neq i} + \delta FRi/Noise \qquad [2.2]$$

Thus, the total change in FR_i is a function of the physical parameter designed to fulfill FR_i (i.e., PPi), the physical parameters fulfilling other FR's and parameters which are unknown to the designer(i.e., noise). For a design to "maintain independence of functional requirements", the last two terms of equation 2.2 must approach zero or be negligible. If these last two terms do not alter FR_i by more that it's specified tolerance, then they can be considered negligible.

For instance, assuming in the auto example of figure 2.1 that the chosen color has a negligible effect on the paints ability to resist water, and that the impermeability properties of the paint do not effect the aesthetics or the color, an uncoupled design matrix can be made as shown in figure 2.3.

Auto Example Design Matrix

$$\begin{Bmatrix} FR_1 \\ FR_2 \end{Bmatrix} = \begin{Bmatrix} \text{Aesthetically Pleasing} \\ \text{Resist Rust} \end{Bmatrix} = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix} \text{X} \begin{Bmatrix} PP_1 = \text{Paint Pigment Color} \\ PP_2 = \text{Paint Permiability} \end{Bmatrix}$$

Figure 2.3 Design matrix for two functional requirements of an automobile.

In this design, the non-diagonal elements A_{12} and A_{21} are approximated at zero, thus making FR_1 and FR_2 uncoupled. Although in the extreme limit one can postulate that all functions are coupled in the physical world by something, if the net effect of all physical coupling is within a tolerable range, then the matrix element can be approximated as zero (i.e., no coupling). The tolerable range or tolerance of the functional requirements must be specified by the designer in order to determine the design matrix.

It thus follows from this argument that the one goal of design should be to minimize the last two terms of equation 2.2 in order to have a more predictable or uncoupled design. The methods developed by Taguchi(1987) are very useful for estimating the effects of parameters and noise on the desired functional requirement.

The basic principle elaborated upon here is that a design which maintains independence of its functional requirements is preferred to one that does not. This is true provided that the uncoupled design is realizable in the process space (i.e., manufacturable or implementable).

2.2.2 Information Axiom

The second principle of design described by Suh(1988) is known as the Information axiom and states:

Minimize the information content of the design

Information is defined here as the knowledge required to fulfill a given set of functional requirements. It follows from this that as a design becomes more complex, the information required to achieve it increases. Thus, to minimize information in the sense meant by the second axiom is to reduce the complexity of a design to a minimum.

The independence axiom is related to the information axiom in that as the complexity of a design increases, the probability of coupling due to additional physical parameters and noise effects is increased. The principle of the information axiom is to reduce this probability and thus increase the certainty that the desired functional requirements are achieved by the chosen physical parameters.

If one considers the range for which the functional requirements are defined as the design range, and the range of possible outcomes for a given set of physical parameters as the system range, then the depiction of the design in figure 2.4 can be used to clarify the meaning of the second axiom.

Design Information

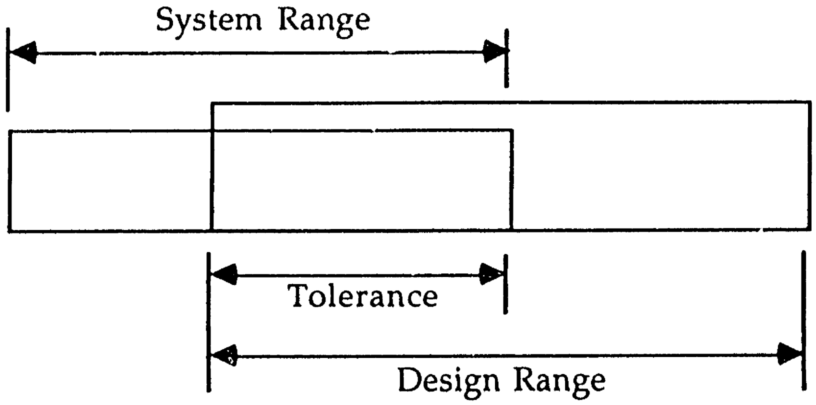


Figure 2.4 The information necessary to achieve a design is proportional to the degree that the design range, defined by the physical parameters, overlaps the system range, defined by the functional requirements.

To minimize information, the tolerance depicted in figure 2.4 as the overlap between the system and design ranges should equal the design range. In this situation when the design range lies completely within the system range, the probability that the functional requirements will be satisfied by the defined physical parameters is one. When the probability that the physical description of a design will satisfy the defined functional requirements is one, then the information required for the design is minimized.

The metric employed by Suh(1988) to quantify the information has its roots in information theory and is the ratio of the design range to the tolerance. The log of this ratio is taken and it is defined as information.

$$I = \text{Log}_2 (\text{Design Range}/ \text{Tolerance}) \qquad [2.3]$$

The problem with this metric is that it does not differentiate between two designs that are both defined completely within the system range. This metric defines both such designs as having zero information content. This is satisfactory in theory when the bounds of the system and design ranges are known with a high degree of certainty. However, in practice it is often found that system ranges vary with changing environmental conditions or noise factors, including unanticipated coupling effects due to other physical parameters as were represented in equation 2.2. Also, the boundaries of the design range are only as precise as the designers perception of the customers requirements and needs, which is a difficult range to define with acuracy. If such fluctuations in the system and design ranges occurs, the two designs that appeared indistinguishable using the metric defined by equation 2.3 may actually be distinguishable, making one more desirable than the other. This situation is illustrated in figure 2.5.

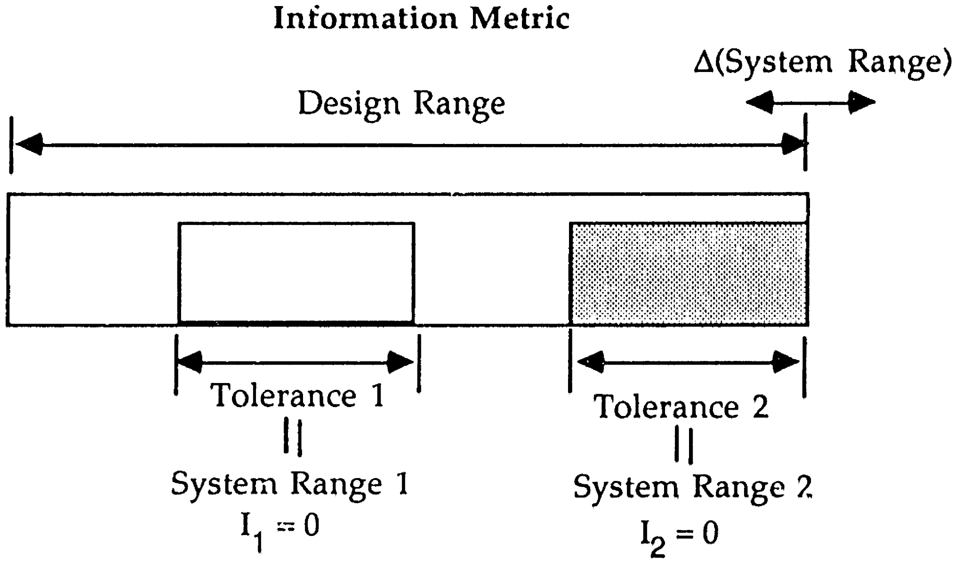


Figure 2.5 Using the information metric described in equation 2.3, both design 1 and 2 above are equally desirable. However, in practice design 1 is more desirable because it is less sensitive to unpredictable fluctuations in system range.

Tightening the tolerance on a design increases the certainty that it will perform as desired. This certainty represents quality or value to the customer. The loss to the customer when a design malfunctions is one measure of this value, which usually does not take the form of a step function. Because the primary goal of a design is to satisfy some customer need, as expressed by the functional requirements, the system ranges in figure 2.5 may be better approximated by another type of function such as a parabola or normal distribution. This step function approximation is correct if one assumes that the customer has no preference within the range of specified functional requirements.

Although this metric for measuring information content may not work in all design contexts, the principle of reducing information content is still sound. To reduce information content is by definition to increase the probability that the defined functional requirements are fulfilled by the chosen physical parameters. The principle should not be confused with a method or metric used in it's application. Thus, one principle may have many different metrics which can be used in it's application, and which depend upon the context of the design.

With this delineation between a method or metric and a fundamental principle or axiom clarified, the application of the independence and information axioms can be further clarified using an example of a passive filter design. In this example, the metrics developed by Taguchi(1987) will be used to apply the design axioms.

2.3 Passive Filter Example

2.3.1 Design Description

The example presented here for comparison has been adopted from Suh(1988) and reworked using the method of engineering design described by Taguchi(1987). Figure 2.6 shows a filter network that must be designed to condition the signal between a Strain-gage Transducer with demodulated output and a Galvanometer.

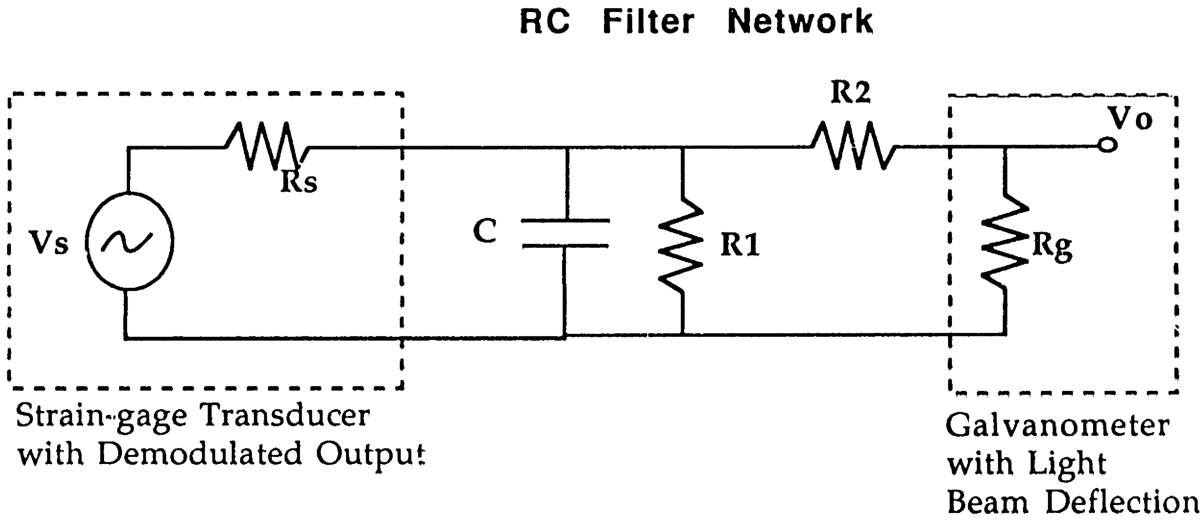


Figure 2.6 A single-pole passive filter network used to condition signal from the Strange-gage Transducer output for use in deflecting the light beam of the Galvanometer.

The Strain-gage Transducer with demodulated output will be modeled as shown in figure 2.6 with an output impedance Rs and a demodulated signal excitation Vs containing the desired strain-gage transducer output and a high-frequency carrier signal. The full-scale Strain-gage output after demodulation will be denoted as V_{fs} [$V_{fs} = (Vs)_{max}$].

The Galvanometer will be modeled as having a terminal impedance Rg and a beam deflection sensitivity Gsen. More detailed descriptions of all the components within the Strain-gage Transducer and Galvanometer are not necessary for the purpose of this example. The characteristic values for the Transducer and Galvanometer just described are given in Table 2.1.

Table 2.1
Characteristic Values and Tolerances
for the Transducer and Galvanometer in figure 2.4.

Characteristic	Nominal Value	Tolerance
Rs (Ω)	120	± 0.15%
V_{fs} (mV)	15	± 0.15%
R_g (Ω)	98	± 0.15%
Gsen (μV/inch)	657.58	± 0.15%

The design problem is being constrained to a single-pole passive filter for which the object becomes choosing the proper set of parameters. Suh(1988) does a spectral analysis of the demodulated outputs that derives an optimal cut-off frequency (Fc) for the filter network that minimizes distortion error due to pass-band error of the transduced output and the error of the non-rejected carrier signal. From this analysis, the minimum error ratio of 12% occurred at a cut-off frequency of 6.84 Hertz, for a single-pole passive filter.

Their are two objectives defined for this passive filter network which can be stated as follows.

Objective 1: Minimize output distortion by placing the filter pole at 6.84 Hz.
Objective 2: Achieve a full-scale beam deflection of ±3 inches by adjusting the filter gain.

The network shown in figure 2.6 can be analyzed using the Kirchoff Current Law to obtain the following transfer function.

$$\frac{V_o}{V_s} = \frac{R_g\ R_1}{(R_2 + R_g)(R_s + R_1) + R_1\ R_s + (R_2 + R_g)R_1\ R_s\ Cs} \qquad [2.4]$$

where s is the Laplace variable. From this transfer function, expressions for the filter cut-off frequency F_c and Galvanometer full-scale Deflection D_{fs} are derived and stated below in equations 2.5 and 2.6.

$$F_c = \frac{R_2\ R_s + R_1\ R_2 + R_g\ R_s + R_g\ R_1 + R_1\ R_s}{2\,\pi\,(R_2 + R_g)\,R_1\,R_s C} \qquad [2.5]$$

$$D_{fs} = \frac{|V_o|}{Gsen} = \frac{|V_{fs}|\, R_g\, R_1}{Gsen(R_2 R_s + R_1 R_2 + R_g R_s + R_g R_1 + R_1 R_s)} \qquad [2.6]$$

The purpose of this example is to illustrate the application of axiomatics rather than designing a complex circuit. However, the same analysis techniques applied to this relatively simple circuit can be carried out on more complex designs as well.

2.3.2 Taguchi Methods

Introduction to Taguchi Methods

Taguchi(1987) has combined traditional engineering methods with statistical methods into an effective way of formalizing information in the design process. The application is meant to reduce uncertainty of the engineered device, product or system and is rooted in the desire to improve the quality of designs.

Taguchi and Phadke(1984) and Clausing(1988) describe the optimization technique that will be employed here for the filter design shown in figure 2.6. This optimization technique consists first of characterizing the parameters of design into four factors and an output response as depicted in figure 2.7.

Taguchi Design Representation

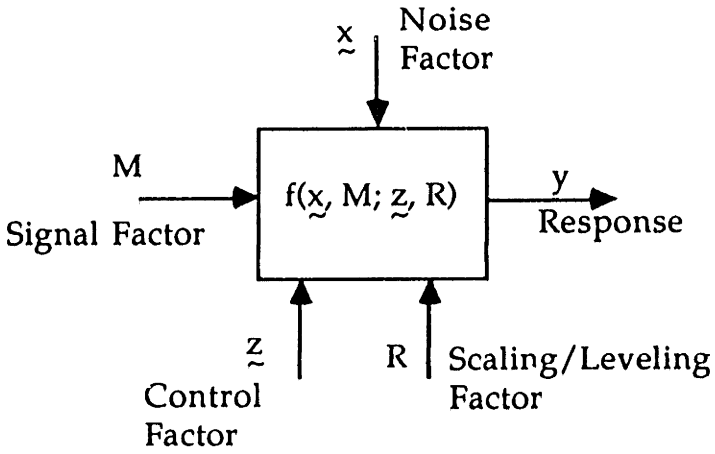


Figure 2.7 A design representation showing the output response y of a design being the function of Noise, Signal, Control and Scaling/Leveling factors.

The response is the output of the design which represents the fulfillment of the defined objectives. A given design may have more than one output response. In the filter example, the output response is characterized as having two components, y_1 and y_2.

$y_1 = F_C$ (cut-off frequency)
$y_2 = D_{fs}$ (full-scale beam deflection)

Signal factors (M) are parameters which can be set by the user of the design to achieve a specific output response. The Strain-gage in the example presented here would represent the signal factor, for which the user may apply various forces to achieve a desired output beam deflection.

Control factors (z) are those physical parameters by which the designer(s) can exercise control over the output response y. In the case of the filter example for which their are two output responses y_1 and y_2, there consequently must be at least two control factors, one for each desired output response. The filter network in figure 2.6 contains three control factors; R_1, R_2 and C. The choosing of control factor values, such that the desired output response is achieved with minimum sensitivity to noise factors, is a primary objective of the Taguchi optimization technique.

Scaling or adjustment factors (R) are special cases of control factors which can be altered to achieve a desired relationship between a signal factor and output response. As can be seen from equations 2.5 and 2.6, any of the three control factors can be used to adjust the cut-off frequency output F_c while only R_1 and R_2 may be used to adjust to the desired beam deflection D_{fs}.

An important assumption in performing this optimization technique is that the chosen control factors should be monotonic in their relationship to the output response. That is, as a control factor is adjusted in one direction, the output response should also either adjust in a constant direction or not at all. If this is not the case, the certainty in predicting the optimum characteristics for a desired response is greatly hindered.

Noise Factors (x) are those variables in a design which are uncontrolable or unpredictable and represent an uncertainty that the desired output response will be achieved. To reduce the effect of this unpredictable component of the output response is the objective of this design method. In the filter example, the tolerances of all the components (i.e, R1, R2, C, Rs, and Rg) and all system characteristics (i.e., Vfs and Gsen) represent the noise factors.

Signal-to-Noise Metric

A metric developed by Taguchi(1987) in order to optimize a design is the ratio of the variation in output response resulting from control factors to that resulting from unpredictable or noise factors. The optimization technique prescribed here says that the better design is the one for which this ratio, which can be called the degree of predictability, is maximized. This qualitative description of the metric used can be described further if one considers any output response to consist of two parts, a predictable and unpredictable part.

That is, for any $y = f(x, M; z, R)$, there exists an equivalent function $y = g(M; z, R) + e(x, M; z, R)$ where g is predictable and e is unpredictable. To optimize y, choose z and R such that the predictability ratio $\Delta g/\Delta e$ is maximized. It is implicitly assumed that z and R are chosen such that the desired response y, based on the design objectives, is achieved. The $\log_{10}$ of this predictability ratio is taken and multiplied by 20 to create a metric which is called the signal-to-noise (SN) ratio η.

$$\eta(x, R) = 20\log_{10}[\Delta g(z, R)/\Delta e(z, R)] \qquad [2.7]$$

The signal-to-noise (SN) ratio is a metric designed to provide a sensitivity measurement of the output response for various levels of control factors and noise. The calculation of this SN ratio is based on the analysis of variance technique which was introduced by Fisher(1925) and others for the purpose of analyzing measurements with several kinds of interdependent effects working simultaneously and determining the relative importance of the various effects. Scheffé(1959) provides a good detailed description of the analysis of variance technique which is worth reading, but is not the main subject of this thesis and therefore is not elaborated upon further here.

For the filter design example presented here, the two output responses F_c and D_{fs} have specified nominal values. The SN ratio for such cases, where a nominal value is the preferred output response, is expressed as a ratio of the mean value μ to the standard deviation σ. The $\log_{10}$ of this ratio is taken and multiplied by 20 to give units of decibels as shown in figure 2.8.

$$\eta = 20\log_{10}E(\mu/\sigma) = 10\log_{10}E(\mu^2/\sigma^2) \qquad [2.8]$$

In equation 2.8, the expected value E of the actual mean squared μ^2 and the standard deviation σ^2 of the output response are used in place of their precise values due to the limited sample size of the experiments. Orthogonal arrays are used to assure a representative sample as explained in the next section.

Since the actual mean is not precisely known, an estimate based on the sample of output responses $y_1, y_2, \ldots y_n$ can be made to approximate μ as described in equation 2.9.

$$E(\mu 2) = (nm^2 - V)/n \qquad [2.9]$$

where m = sample mean = $(\Sigma_n y)/n$
V = total error variation of y over the entire range of n samples
n = number of samples

The error variation V is used to represent the expected value of the standard deviation squared. This is an estimate of the unpredictable part of the output response depicted in equation 2.7. This variation is calculated from the sample responses as shown below in equation 2.10.

$$E(\sigma^2) = V = (\Sigma y^2 - S_m)/(n-1) \qquad [2.10]$$

where $S_m = (\Sigma y)^2/n$

Combining equations 2.8, 2.9 and 2.10, the SN ratio as a function of the output response samples $y_1, y_2, \ldots y_n$ can be expressed as shown in figure 2.11.

$$\eta = 10\log_{10} E(\mu^2/\sigma^2) = 10\log_{10}[(S_m-V)/nV] \qquad [2.11]$$

The accuracy of this statistical technique is dependent on the sampled output response as stated earlier. It is necessary to use a sample of responses that is large enough to fully represent the broad range of possible outcomes while not being so large that the experimentation becomes overly time consuming and impractical. Taguchi(1987) suggests the use of orthogonal arrays for this purpose.

Orthogonal Arrays

The origin of orthogonal arrays goes back to Euler's Graeco-Latin squares and the use of orthogonal arrays for the design of experiments was first studied during World War II. Figure 2.8 shows an orthogonal array with 9 rows of experiments denoted by L_9.

Origin of Orthogonal Arrays

L_9

Exp	A	B	C	D
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

L_1

1	2	3
2	3	1
3	1	2

L_2

1	3	2
2	1	3
3	2	1

Figure 2.8 An L_9 orthogonal array and two 3X3 Latin squares denoted by L_1 and L_2 which were used to construct the L_9 array.

The L_9 orthogonal array is constructed from the Latin squares L_1 and L_2 shown in figure 2.8. For each of the four factors A, B, C, and D of the L_9 there are three levels which are represented in the columns. Each of the nine rows represents one combination of factor levels or one experiment.

There are nine combinations of levels between any two columns of the L_9, (11), (12), (13), (21), (22), (23), (31), (32) and (33) for which each combination appears with the same frequency. Any two such columns are said to be balanced or orthogonal.

Orthogonal arrays are used in design of experiments in order to find the average effect of a factor as all other factors are varied. For example, factor A has three levels (i.e., A_1, A_2 and A_3). When the influence of A_1, A_2 and A_3 on the outcome is large this influence will appear great, even when the conditions of the other factors are varied. When this influence changes as different combinations of factors are used, the effect of A will appear small. When orthogonal arrays are used, the estimates of the main effects of all factors A, B, C, and D are independent assuming normality and equality of error variance.

The main effect of factors can be estimated by summing the squares of the effects at various levels and averaging over the number of degrees of freedom at each level. For example, the main effect of A on the variation in output response $y_1, y_2, \ldots y_9$ obtained through experimenting with the nine combinations of factors shown in the L_9 orthogonal array of figure 2.8 can be calculated as follows.

The effect of A on the total variation in the output y equals the total variation in the output while A is at level 1 [i.e., $(y_1 - m) + (y_2 - m) + (y_3 - m)$], plus the output variation while A is at level 2, plus the output variation while A is at level 3. These output variations are estimated as the sum of the difference between the output $y_1, y_2, \ldots y_9$ from the sample mean m for the outputs corresponding to the same level of A. Each of these variations is squared to represent the effect due to A for the variation in output response at the respective levels of A. These squared sums are divided by their respective degrees of freedom and summed to give a measure for the total effect of A on the variation in output response y.

This sum-of-the-squares calculation measuring the effect of A for the nine outputs $y_1, y_2, \ldots y_9$ corresponding to the nine experiments in figure 2.8 is shown in equation 2.12.

$$\begin{aligned} \text{Effect of A} = \; & [(y_1 - m) + (y_2 - m) + (y_3 - m)]^2/2 \\ & + [(y_4 - m) + (y_5 - m) + (y_6 - m)]^2/2 \\ & + [(y_7 - m) + (y_8 - m) + (y_9 - m)]^2/2 \end{aligned} \qquad [2.12]$$

where m = the sample mean = $(\Sigma_n y)/n$

In the circuit example to follow, this measurement of the effect of the various factors described here will be calculated for both the SN and sample mean m of the output response. These measures will be used to help optimize the design by minimizing interactions between factors and by minimizing the effects of noise on the desired output response.

Experiment

The first step in the analysis of the filter circuit is to choose a control factor to adjust each desired output response to its desired performance objective. An analysis of parameters is done using orthogonal arrays to measure the relative influence of each control factor on the output response. Control factors that are both insensitive to noise factors(i.e., large η) over a broad range of values and have a linear relationship to the output response make desirable adjustment factors. In the case of the filter design presented here, two adjustment factors are required, one to adjust the frequency response F_C and one to adjust the maximum light beam deflection D_{fs}. In choosing the adjustment factors, the first axiom will be applied to the results provided by the Taguchi analysis by choosing adjustment factors that maintain the independence of the functional requirements.

Other control factors which are not used as adjustment factors will be used to maximize the SN ratio. By experimenting with different levels of control factors, the levels that minimize the desired output responses sensitivity to noise is sought. Minimizing the output sensitivity to noise (i.e., maximizing SN) is analogous to tightening the design range illustrated in figure 2.4 and is thus applying the second axiom of minimizing information content.

It is desirable to check for noise insensitivity and linear dependence over the entire range of possible control factor values. Orthogonal arrays are used to limit the various combinations of control factor values to a finite and reasonable number, while choosing a sample that is representative of the entire range of possible combinations. For the three control factors of the filter example, a low, medium and high level or value of control factor is chosen and are shown in Table 2.2.

Table 2.2
Levels for Control Factors

Level	1	2	3
R1 (Ω)	20	50000	100000
R2 (Ω)	0.01	265	525
C (μf)	1400	815	231

To choose the levels of control factors given in Table 2.2, the broadest range of values that are capable of fulfilling the desired design objectives should be selected. Choosing levels for which the preferred output response becomes impossible may yield impractical results and should thus be avoided. Conversely, choosing levels within too narrow of a range may not yield an optimal design if the optimal lies outside of the range of levels chosen. Thus, it is important to choose a range of control factor levels that is representative of the range of practical possibilities and in general it is better to choose too broad of a range than too narrow.

To test all possible combinations of levels for the three factors in Table 2.2, it would take $3^3=27$ experiments. An L_9 orthogonal array can be employed in this example to reduce the number of experiments from 27 to only 9 combinations of control factors as shown in figure 2.8. Since there are only three factors in this example, the first column of the L_9 orthogonal array shown in figure 2.8 has been omitted and the resulting array of combinations for the three control factors, each having three levels is shown in Table 2.3.

Table 2.3
Combinations of Control Factor Levels
in an L9 Orthogonal Array

Experimen	R1 (Ω)	R2 (Ω)	C (µf)
1	20	0.01	1400
2	50000	265	815
3	100000	525	231
4	20	265	231
5	50000	525	1400
6	100000	0.01	815
7	20	525	815
8	50000	0.01	231
9	100000	265	1400

For each of the nine combinations of control factor levels shown above in Table 2.3, two calculations will be made for each desired output response (i.e., F_C and D_{fs}). A second orthogonal array will be defined for each row in the L_9 orthogonal array of table 2.3 as the outer array. The purpose of this outer array will be to measure the variation in output response due to variation in the tolerance of the control factors. Each of these experiments will yield a mean value m and a SN value η for each output response as calculated in equations 2.9 and 2.11 respectively.

The analysis of variation in output due to noise factors is carried out employing the SN metric described in equation 2.11. The SN metric will give an indication of how much the desired output response changes as the tolerances for the control factors are varied over their respective ranges. There are seven noise factors in the filter example being used here as previously indicated and three levels are chosen for each factor at their high, medium and low tolerance limits as shown in Table 2.4.

Table 2.4
Levels for Noise Factors

Factor	Levels		
	1	2	3
R1 (Ω)	R1 - 5%	R1	R1 + 5%
R2 (Ω)	R2 - 5%	R2	R2 + 5%
C (μf)	C - 5%	C	C + 5%
Rs (Ω)	119.82	120	120.18
Rg (Ω)	97.853	98	98.147
Gsen (μV/inch)	656.594	657.58	658.566
Vd (Volts)	0.014978	0.015	0.015023

The nominal value for the control factors R_1, R_2, and C in Table 2.4 will be determined by the current experiment being conducted. The nominal values for each experiment are indicated in Table 2.3. For example, the values of R_1, R_2, and C for the first experiment will be 20Ω, 0.01Ω and 1400μf as indicated in the first row of Table 2.3. The goal of the first, as well as each subsequent experiment, is to determine how the output response varies as a given combination of levels (i.e., 20Ω, 0.01Ω and 1400μf) are pushed to the edge of their tolerances or noise levels.

As shown in Table 2.4, there are seven noise factors with three levels each, for which an exhaustive test of every combination possible would require 3^7=2187 experiments. As described earlier, the use of orthogonal arrays can help minimize the number of combinations necessary while maintaining a representative sample. Since their are seven control factors with three levels each, the L_9 orthogonal array described earlier will not be sufficient. An L_{27} orthogonal array can accommodate as many as 13 control factors of three levels each, which will be more than sufficient for this example.

An L_{27} orthogonal array may be constructed from the L_9 array of figure 2.8 as shown in figure 2.9. As with the L_9 array, the nine combinations of levels between any two columns of the L_{27}, (11), (12), (13), (21), (22), (23), (31), (32) and (33) appears with the same frequency. Therefore, any two such columns are said to be balanced or orthogonal.

Construction of an L_{27} Orthogonal Array

Exp.	A	B	C	D	E	F	G	H	I	J	K	L	M
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	2	2	2	2	2	2	2	2	2
3	1	1	1	1	3	3	3	3	3	3	3	3	3
4	1	2	2	2	1	1	1	2	2	2	3	3	3
5	1	2	2	2	2	2	2	3	3	3	1	1	1
6	1	2	2	2	3	3	3	1	1	1	2	2	2
7	1	3	3	3	1	1	1	3	3	3	2	2	2
8	1	3	3	3	2	2	2	1	1	1	3	3	3
9	1	3	3	3	3	3	3	2	2	2	1	1	1
10	2	1	2	3	1	2	3	1	2	3	1	2	3
11	2	1	2	3	2	3	1	2	3	1	2	3	1
12	2	1	2	3	3	1	2	3	1	2	3	1	2
13	2	2	3	1	1	2	3	2	3	1	3	1	2
14	2	2	3	1	2	3	1	3	1	2	1	2	3
15	2	2	3	1	3	1	2	1	2	3	2	3	1
16	2	3	1	2	1	2	3	3	1	2	2	3	1
17	2	3	1	2	2	3	1	1	2	3	3	1	2
18	2	3	1	2	3	1	2	2	3	1	1	2	3
19	3	**1**	3	2	**1**	3	2	**1**	3	2	**1**	3	2
20	3	**1**	3	2	**2**	1	3	**2**	1	3	**2**	1	3
21	3	**1**	3	2	**3**	2	1	**3**	2	1	**3**	2	1
22	3	**2**	1	3	**1**	3	2	**2**	1	3	**3**	2	1
23	3	**2**	1	3	**2**	1	3	**3**	2	1	**1**	3	2
24	3	**2**	1	3	**3**	2	1	**1**	3	2	**2**	1	3
25	3	**3**	2	1	**1**	3	2	**3**	2	1	**2**	1	3
26	3	**3**	2	1	**2**	1	3	**1**	3	2	**3**	2	1
27	3	**3**	2	1	**3**	2	1	**2**	1	3	**1**	3	2

Figure 2.9 An L_{27} orthogonal array can be used to experiment with up to 13 factors (A, B, . . . M) each having 3 levels. The bold type shows how the L_9 orthogonal array of figure 2.8 was used in constructing the L_{27} depicted here.

The L_{27} orthogonal array is used in this example to form an outer array for each experiment (i.e., combination of control factors) shown in the L_9 array of table 2.3. Table 2.5 shows the L_{27} outer array for the first experiment of table 2.3 from which the output responses F_c and D_{fs} are calculated for each of the 27 combinations of noise levels, using equations 2.5 and 2.6 respectively. Because there are only seven noise factors in this example and the L_{27} in figure 2.9 can accommodate thirteen factors, only the last seven columns are utilized in Table 2.5.

Table 2.5
27 Combinations of Noise Factors
Tested for Experiment 1 values from Table 2.3
with the Output Responses Calculated in the last two columns

Noise Factor Combinations								Outputs	
Data No.	R1 (Ω)	R2 (Ω) (Ω)	C (μf)	Rg (Ω)	Rs (Ω)	Gsen (μV/inch)	Vs (mV)	Dfs (inches)	Fc (Hz)
1	21	0.0105	1470	84.42	120.18	566.4582	15.075	3.266	7.339
2	20	0.01	1400	84	120	563.64	15	3.157	7.985
3	19	0.0095	1330	83.58	119.82	560.8218	14.925	3.045	8.728
4	21	0.01	1400	84	119.82	560.8218	14.925	3.272	7.715
5	20	0.0095	1330	83.58	120.18	566.4582	15.075	3.150	8.411
6	19	0.0105	1470	84.42	120	563.64	15	3.046	7.883
7	21	0.0095	1330	83.58	120	563.64	15	3.265	8.127
8	20	0.0105	1470	84.42	119.82	560.8218	14.925	3.164	7.599
9	19	0.01	1400	84	120.18	566.4582	15.075	3.039	8.282
10	19	0.0105	1400	83.58	120.18	563.64	14.925	3.021	8.289
11	21	0.01	1330	84.42	120	560.8218	15.075	3.304	8.113
12	20	0.0095	1470	84	119.82	566.4582	15	3.146	7.606
13	19	0.01	1330	84.42	119.82	566.4582	15	3.034	8.714
14	21	0.0095	1470	84	120.18	563.64	14.925	3.247	7.345
15	20	0.0105	1400	83.58	120	560.8218	15.075	3.186	7.991
16	19	0.0095	1470	84	120	560.8218	15.075	3.074	7.889
17	21	0.0105	1400	83.58	119.82	566.4582	15	3.253	7.722
18	20	0.01	1330	84.42	120.18	563.64	14.925	3.140	8.396
19	20	0.0105	1330	84	120.18	560.8218	15	3.169	8.403
20	19	0.01	1470	83.58	120	566.4582	14.925	3.010	7.896
21	21	0.0095	1400	84.42	119.82	563.64	15.075	3.291	7.709
22	20	0.01	1470	83.58	119.82	563.64	15.075	3.174	7.612
23	19	0.0095	1400	84.42	120.18	560.8218	15	3.057	8.276
24	21	0.0105	1330	84	120	566.4582	14.925	3.235	8.120
25	20	0.0095	1400	84.42	120	566.4582	14.925	3.128	7.978
26	19	0.0105	1330	84	119.82	563.64	15.075	3.062	8.721
27	21	0.01	1470	83.58	120.18	560.8218	15	3.277	7.352

Results

From the output data in Table 2.5, the SN ratio value for each output response (i.e., F_C and D_{FS}) can be calculated from equation 2.11. The sample mean value for each output response is also calculated over the various combinations of noises by summing the outputs and dividing by the number of data points. This procedure is repeated for each of the nine experiments in Table 2.3 and the SN ratio η and median output response m is reported below in Table 2.6 for each output response.

Table 2.6
Signal-to-Noise and mean output response for each control factor combination given in Table 2.3.

Experiment	Fc		Dfs	
	SN (dB)	m (Hz)	SN (dB)	m (inches)
1	25.821	8.008	30.388	3.156
2	27.443	2.195	32.218	4.761
3	27.517	6.893	30.234	3.066
4	25.378	42.295	26.756	0.873
5	27.517	1.138	30.233	3.063
6	27.591	3.960	43.857	10.952
7	25.309	11.748	26.069	0.511
8	27.591	13.980	43.858	10.947
9	27.443	1.277	32.220	4.765

The SN results given in Table 2.6 represent a measure of output response sensitivity to noise for nine combinations of control factor levels. The sample mean m is an indication of the variation in the output response over the nine different combinations of control factor levels. From the results in table 2.6, an indication of the relative influence that each of the control factors had on both the SN ratio and the mean value for each output response can be calculated.

For example, the influence that control factor R_1 had on the output sensitivity to noise (i.e., SN ratio) can be calculated by taking the average SN ratio for R_1 ($\pm \Delta R_1$) at each level (i.e., 20Ω, 50KΩ and 100KΩ), to obtain three data points. These data points can also be calculated for the other control factors and compared for relative magnitudes and variations over the range of levels.

Similarly, the relative influence of each control factor on the mean value can be calculated by looking at an average value for the sample mean m at each level of control factor. Control factors showing a constant SN response over the entire range of levels while showing a linear relationship to the mean output response are desirable adjustment factors as was previously mentioned.

A measure of the influence of each control factor on each output response can be measured using the summation of squares method described in equation 2.12. For example, to determine the summation of squares SR_1 for the SN ratio of the factor R_1 and response F_C, the results in the first column of table 2.6 (indicated by $SN_1, SN_2, \ldots SN_9$) will be used as shown below in equation 2.13.

$$\begin{aligned} SR_1 = \; & 1/2[(SN_1 - m_{sn}) + (SN_4 - m_{sn}) + (SN_7 - m_{sn})]^2 \\ & + 1/2[(SN_2 - m_{sn}) + (SN_5 - m_{sn}) + (SN_8 - m_{sn})]^2 \\ & + 1/2[(SN_3 - m_{sn}) + (SN_6 - m_{sn}) + (SN_9 - m_{sn})]^2 - CF \end{aligned} \quad [2.13]$$

where m_{sn} is the sample mean for the calculated SN values in column one of table 2.6. The correction factor CF is defined as $[\Sigma_9(SN_n - m_{sn})]^2/9$.

These calculations are repeated for each control factor with both the SN ratio and the mean response of each output and the results are reported in Tables 2.7a, 2.7b, 2.7c, and 2.7d.

Table 2.3 shows that level 1 of R_1 was used during experiments 1, 4 and 7 which yielded SN ratios for the cut-off frequency output F_C of 25.821, 25.378 and 25.309, respectively, as shown in Table 2.6. The mean of these three SN ratios is taken to yield an average output sensitivity for R_1 at level 1 of 25.503 as indicated in Table 2.7a. Similarly, the mean values for F_C obtained while R1 is at level 1 are calculated by averaging the respective data points in Table 2.3 which yielded an average of 20.684 Hertz for F_C, level 1 as indicated in Table 2.7b. This is repeated for all three control factors and for both output responses, and the results are presented in Tables 2.7a, 2.7b, 2.7c, and 2.7d.

Table 2.7a

Analysis of Variance for the SN of F_C

Factor	Level Means 1	2	3	Sum of Squares
R_1	25.503	27.517	27.517	12.170
R_2	27.001	26.755	26.781	0.164
C	26.927	26.781	26.829	0.050

Table 2.7b

Analysis of Variance for the Mean of F_c

Factor	Level Means			Sum of Squares
	1	2	3	
R_1	20.684	5.771	4.043	753.402
R_2	8.649	15.256	6.593	184.358
C	3.474	5.968	21.056	814.464

Table 2.7c

Analysis of Variance for the SN of D_{fs}

Factor	Level Means			Sum of Squares
	1	2	3	
R_1	27.738	35.436	35.437	177.818
R_2	39.368	30.398	28.846	290.378
C	30.947	34.048	33.616	25.392

Table 2.7d

Analysis of Variance for the Mean of D_{fs}

Factor	Level Means			Sum of Squares
	1	2	3	
R_1	1.513	6.257	6.261	67.569
R_2	8.352	3.466	2.213	94.684
C	3.661	5.408	4.962	7.411

Using the Independence Axiom to Choose Adjustment Factors

These results can be used to apply the independence axiom in the selection of adjustment factors. As explained earlier, control factors that are both insensitive to noise factors(i.e., large and constant η) over a broad range of values and have a linear relationship to the output response make desirable adjustment factors. In addition, for designs with two or more output objectives that must be optimized (e.g., Fc and Dfs), the independence of these outputs relative to their respective adjustment factors must be maintained.

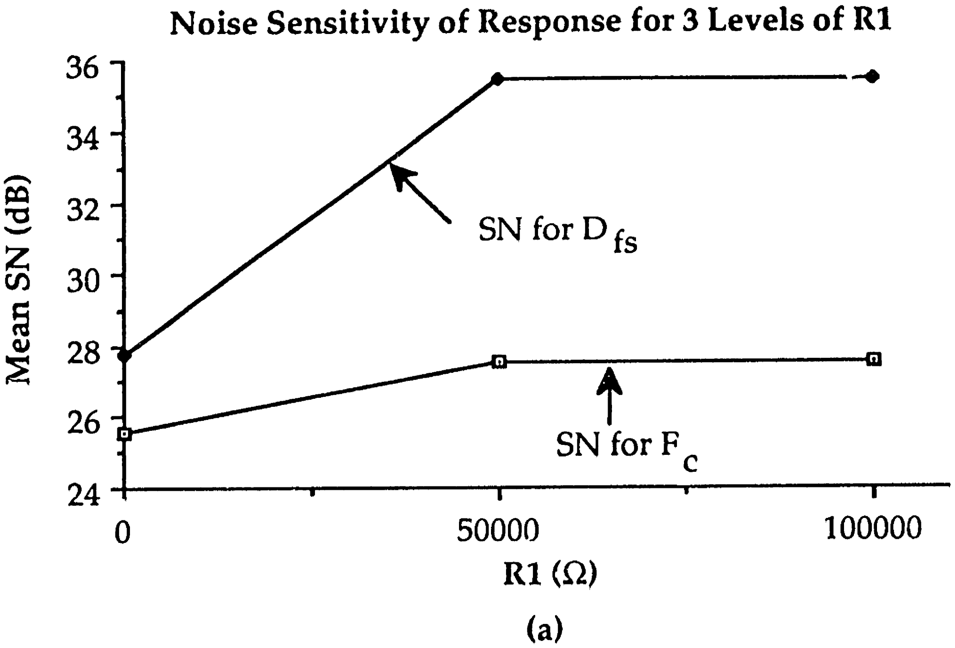
For example, based on the independence axiom, the control factor with the most significant contribution to the mean response of Fc, while contributing a minimum to the mean response of D_{fs}, should be chosen as an adjustment factor for F_c. The control factor that best fits this description is clearly C as indicated by its minimal contribution to the sum of squares in Table 2.7d and its significant contribution to the sum of squares in table 2.7b. Thus, the sum of squares metric can be used as an approximation of the coupling of functional requirements by the various physical parameters or control factors.

It can also be helpful to plot the results in Tables 2.7a, 2.7b, 2.7c, and 2.7d in order to gain some two dimensional insight as to how the control factors are related over their entire range of values to the output responses mean value and SN ratio. These plots have been done and are shown in figures 2.10, 2.11 and 2.12.

The Graphs in figures 2.12a and 2.12b indicate that C is a good choice for an adjustment factor due to its relative constant SN performance over the broad range of its level values. The mean output response for F_C does not vary exactly linear with respect to C, but does have a monotonic and almost linear relationship. For these reasons, C is the most appropriate choice of adjustment factor for the frequency output response F_C.

Similarly, the results in Table 2.7b indicate that the contribution to the variation of the cut-off frequency response Fc is significantly greater for R_1 than it is for R_2. Also, the contribution of R_2 to the variation in the mean response of Dfs is greater than the contribution of R_1 as indicated in Table 2.7d. This indicates that R_2 is coupled by a lesser degree to the output Fc than R_1, while also contributing more toward the variation in the deflection output D_{fs}. Therefore, based on the independence axiom, R_2 would make a better adjustment factor than R_1 for the deflection output response D_{fs} of this filter network.

The results of this Taguchi analysis have been used to aid in the selection of adjustment factors or physical parameters that form a design with the least coupling. This is the same as choosing physical parameters that maintain the independence of functional requirements. Notice from the sum of the squares column of table 2.7b that adjustment factor R_2 does influence the frequency output response to some degree and thus the design is not totally uncoupled. With R_2 and C chosen as adjustment factors, the optimization and the consequent selection of parameter values can be done using the SN metric and applying the information axiom.



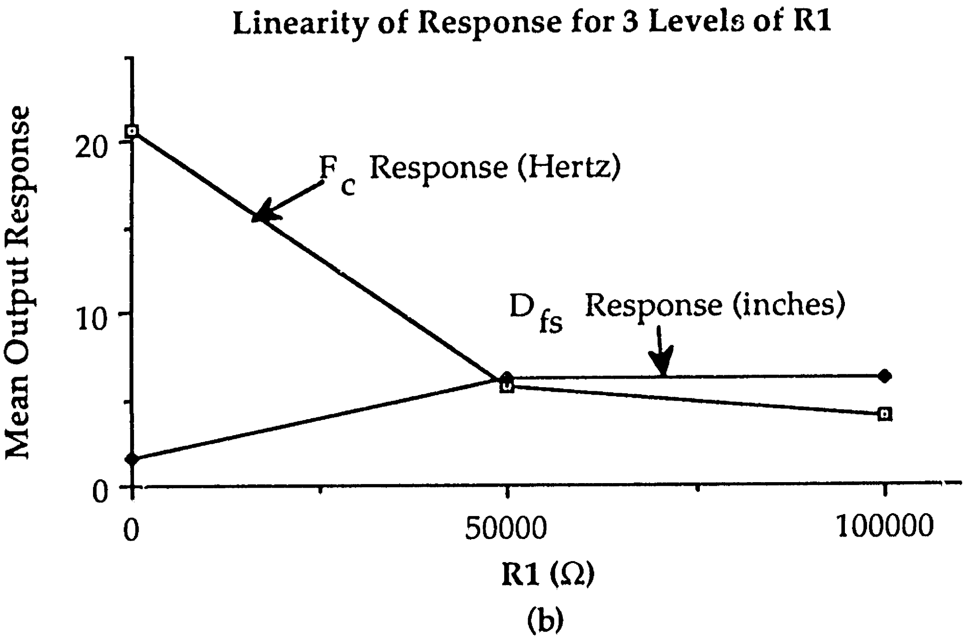


Figure 2.10 (a) The variation in output response sensitivity(i.e., SN ratio) as R_1 is varied over three values from 20-100kΩ. (b) The mean output response measured as R_1 is varied over three levels.

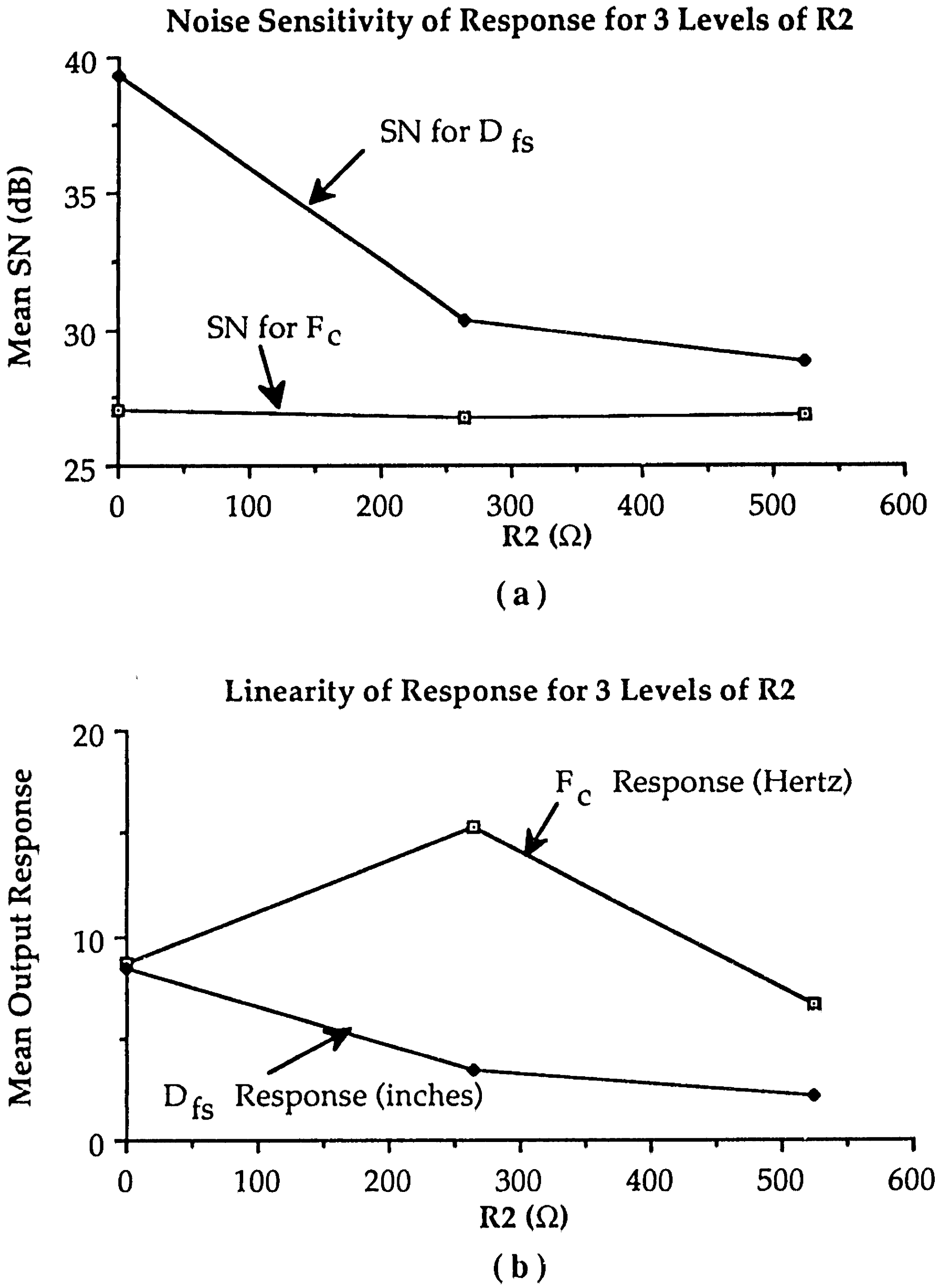


Figure 2.11 (a) The variation in output response sensitivity(i.e., SN ratio) as R_2 is varied over three values from .01-525Ω. (b) The mean output response measured as R_2 is varied over three levels.

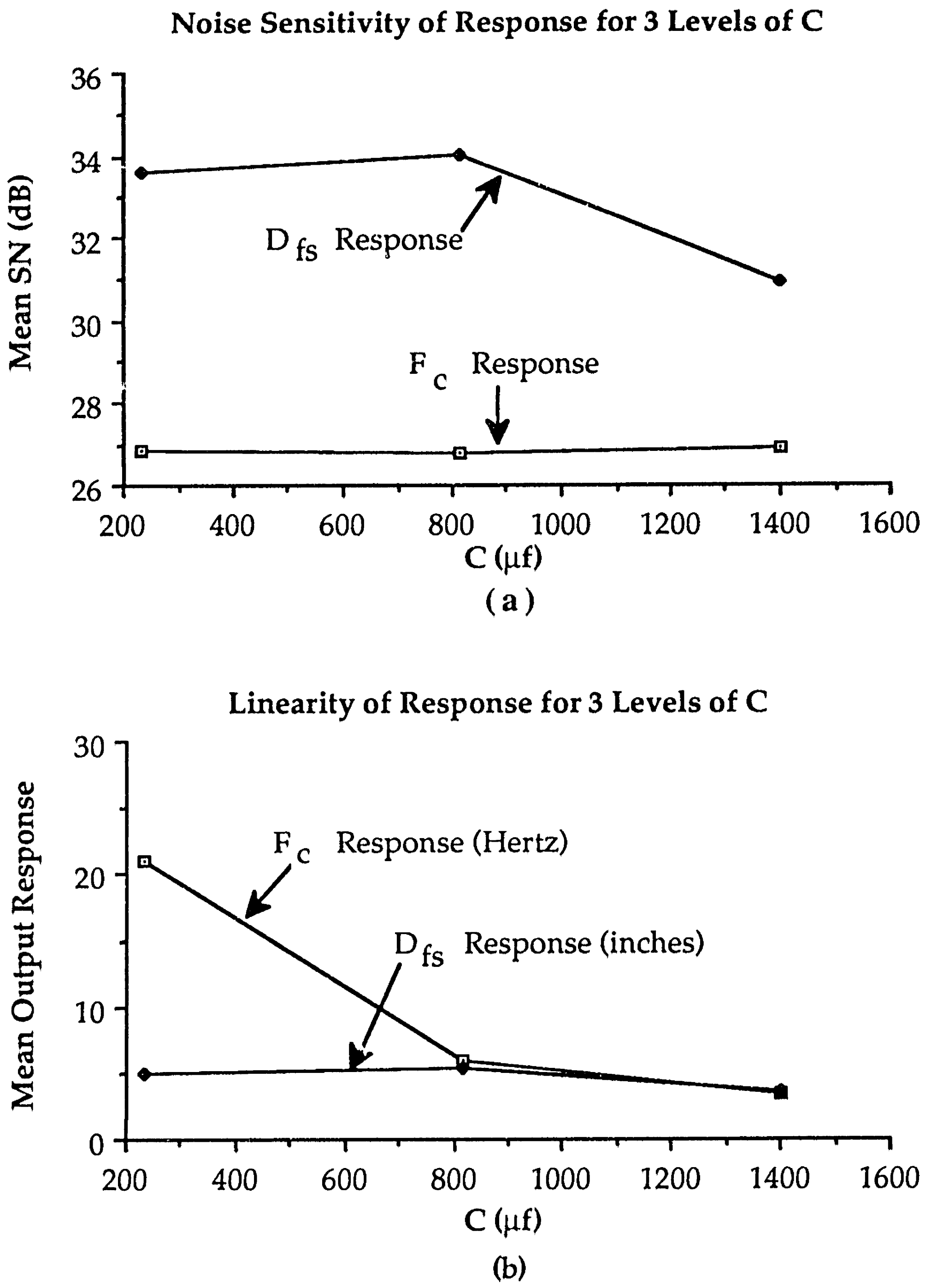


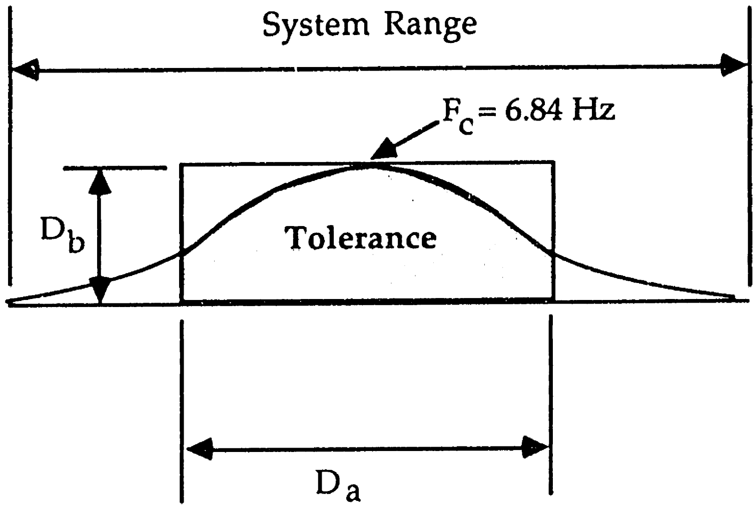
Figure 2.12 (a) The variation in output response sensitivity(i.e., SN ratio) as C is varied over three values from 1400-231μf. (b) The mean output response measured as C is varied over three levels.

Using the Information Axiom to Optimize

To minimize information in the context of this example is analogous to tightening the system range of figure 2.4 so that it is completely within the design range. In this filter example, there are two system ranges, one for each output response Fc and Dfs. The allowable fluctuation of these outputs as defined by the design specifications constitutes the design range.

The design range should be a reflection of the value the customer attributes to the design at different levels of the output response. For example, the desired system response for F_c was defined to be 6.84 Hertz which has the most value to the customer. A little less or a little more than this target of 6.84 Hertz would represent a loss to the value of the design from the perspective of the customer. A tolerance for the output response is determined by deciding how much variation in F_c would constitute an acceptable range to the majority of customers. Customer value, which is a function of the ability of the system range, is usually not a step function and is more likely to be approximated by a continuous function as illustrated in figure 2.13.

A Continuous System Range



Design Range = $D_a \times D_b$

Figure 2.13 The system range is for the frequency output response Fc is approximated with a continuous function. As the system range is tightened around the target value of 6.84 Hertz, ratio of the design range to the tolerance is decreased.

Typically there is a trade-off to be made between how much information should be required of the customer to provide versus the cost of the design. For example, as the component tolerances in the filter circuit are tightened, the system range is tightened around the target value of 6.84 Hertz and the information content, defined in equation 2.3 as the ratio of the design range to the tolerance, is subsequently decreased. As components with tighter tolerances are required, cost will tend to increase. How to make this trade-off between the information content of a design and its cost is not the subject of this analysis, but rather how to measure information content that will aid in making such trade-off decisions.

The concept of tolerance design and reducing information content are the same in the context of this filter example. The information content of the filter design can be reduced by tightening the tolerances of the components R_1, R_2 and C. The SN ratio is a measure of the variation in the output response as the levels (i.e., noise levels) of the noise factors shown in table 2.5 are varied over their respective range of tolerances which are defined in table

2.4. The SN ratio provides a measure of the system range for specified set of control factor values and tolerances.

By observing the sum of the squares column of tables 2.7a and 2.7b, the control factor that contributes the most to variations in the SN ratio can be seen. For example, table 2.7a indicates that control factor R_1 provides the greatest contribution to variations in the SN ratio. This variation in SN ratio translates into fluctuations in the system range and consequently an unpredictable design. One way to minimize this unpredictability is to tighten the tolerance of the control factor R_1 which would reduce the variation in SN ratio contributed by R_1. Using components with tighter tolerances often results in higher costs.

A second way to minimize the information content of the filter design without changing the tolerance of the components is to adjust the level of the non-adjustment control factors (i.e., R_1) over a broad range of possible values while using the adjustment factors (i.e., R_2 and C) to adjust the output responses (i.e., D_{fs} and F_c) to their target values. Then using these values and the outer orthogonal array of figure 2.5, calculate the SN ratio for both output responses F_c and D_{fs}. The set of values that maximizes the SN ratio for both output responses will minimize the system range and is thus the optimal design based on the information axiom. This technique for minimizing information content using the SN ratio was done for the filter design and the results are reported in table 2.8.

Table 2.8

Signal-to-noise Optimization with R2 and C as the adjustment factors

Trial	Control Factor Values			Signal-to-Noise Ratio			Target Values	
	R1 (Ω)	R2 (Ω)	C (μf)	Dfs (dB)	Fc (dB)	Total (dB)	Dfs (inches)	Fc (Hertz)
1	23	2.55	1436.98	30.6880	25.9259	56.613897	3	6.84
2	100	186.16	508.47	30.1266	26.7419	56.868488	3	6.84
3	500	406.16	286.59	30.3496	27.4136	57.763165	3	6.84
4	1000	460.18	258.86	30.3300	27.4865	57.816475	3	6.84
5	**1060**	**463.6**	**257.29**	**30.327**	**27.49**	**57.81658**	**3**	**6.84**
6	1125	466.90	255.78	30.3237	27.4926	57.816302	3	6.84
7	1250	472.4	253.31	30.3179	27.497	57.814999	3	6.84
8	1500	480.85	249.61	30.3078	27.503	57.810942	3	6.84
9	2000	491.77	244.99	30.2926	27.5094	57.801983	3	6.84
10	10000	519.74	233.90	30.2426	27.5172	57.759703	3	6.84
11	100000	526.41	231.40	30.2283	27.5172	57.745503	3	6.84
12	1000000	527.08	231.15	30.2268	27.5172	57.743978	3	6.84
13	1000000000	527.16	231.12	30.2266	27.5172	57.743807	3	6.84

Since there are two output responses (i.e., FRs) and consequently two SN ratios, the optimization strategy employed here is to treat both outputs equally and thus maximize the sum of the two SN ratios. This may not necessarily be desired in all situations. For example, if the system range of one output response is significantly greater than that of a second, it may be more desirable to maximize the SN ratio of the second response than that of the first. The goal of this strategy is to pick parameter values that will achieve the functional requirements and maximize the SN ratio.

When the SN ratio is maximized, the system range is tightened and the information content of the design is consequently minimized. The results shown in Table 2.8 indicated that information is minimized(i.e., Total SN is maximized) for the parameter values of R1=1060Ω, R2=464Ω, and C=257μf.

Note that if the strategy were to maximize the SN for the Fc response and ignore the Dfs response, then the optimal network would be achieved as R1 approaches a very large value. This is equivalent to removing the parallel resistor R_1 completely from the circuit.

The SN ratio was employed here as an indication of the size of the system range relative to the size of the design range described by figure 2.13. Maximizing the SN ratio serves to reduce variations in the system range around the target output response values. Reducing this variation (i.e., uncertainty due to noise factors) in the system range about the targeted output response has the effect of increasing the tolerance shown in figure 2.14.

Maximizing the Signal-to-Noise Ratio
by Tightening the System Range Tolerance

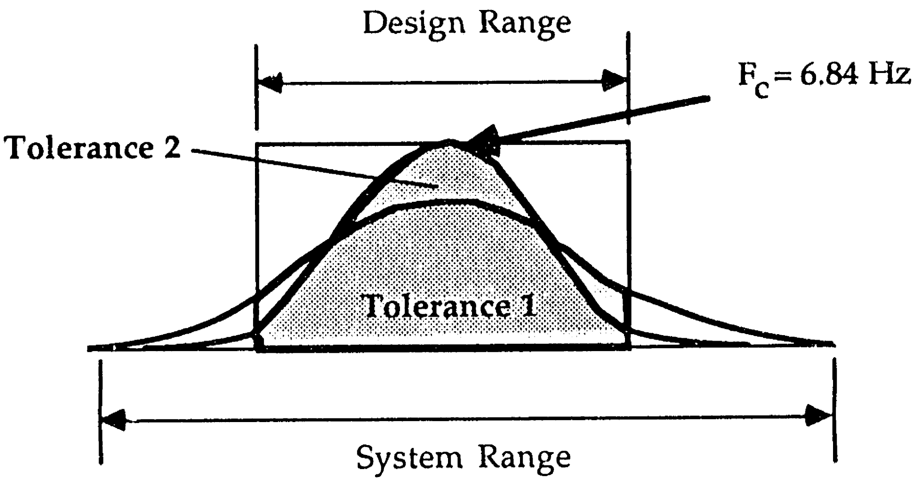


Figure 2.14 As the signal-to-noise ratio is maximized, the variation in the system range is decreased. When the system range is tightened about the target output value for F_C, the tolerance is consequently increased (e.g., Tolerance 2 > Tolerance 1) and information content is decreased.

Therefore, from equation 2.3, an increase in the tolerance while maintaining a constant design range is equivalent to decreasing the information content. Thus, by maximizing the SN ratio about the targeted response, the information content of the design is minimized. In this way the SN metric developed by Taguchi can be used to apply the information axiom towards the optimization of a design.

2.3.3 Another Method for Measuring Functional Coupling

The circuit in figure 2.6 can be analyzed by forming a design matrix as was illustrated in figure 2.2. By choosing the physical parameter C to adjust F_C and R_2 to adjust D_{fs}, the matrix can be defined as shown in figure 2.15.

RC Filter Design Matrix

$$\left\{ \begin{matrix} F_c \\ D_{fs} \end{matrix} \right\} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \mathrm{X} \left\{ \begin{matrix} C \\ R_2 \end{matrix} \right\}$$

Figure 2.15 A design matrix representation for the RC Filter shown in figure 2.6.

The matrix elements can be calculated by applying equation 2.1 to the analytical expressions describing the filter network in equations 2.5 and 2.6 as follows.

$$A_{11} = (\partial Fc/\partial C)/Fc = -1/C$$

$$A_{12} = (\partial Fc/\partial R_2)/Fc = -1/(R_2+Rg) + \{(Rg+R_2)/[Rs(R_2+R_1+Rg) + R_1(R_2+Rg)]\}$$

$$A_{21} = (\partial D_{fs}/\partial C)/D_{fs} = 0$$

$$A_{22} = (\partial Fc/\partial R_2)/D_{fs} = -(Rs+R_1)/(R_2Rs+R1R_2+RgRs+RgR_1+R_1Rs)$$

If the values of R_1, R_2 and C are chosen as those that were optimized by the application of Taguchi methods in the previous section to this design matrix, the design matrix can be numerically represented as shown in Figure 2.16.

RC Filter Design Matrix Values

$R_1 = 1060\Omega$, $R_2 = 463\Omega$, $C = 257\mu f$

$$\left\{ \begin{matrix} F_c \\ D_{fs} \end{matrix} \right\} = \begin{bmatrix} -3891 & -0.0003152 \\ 0 & -0.0014399 \end{bmatrix} \mathrm{X} \left\{ \begin{matrix} C \\ R_2 \end{matrix} \right\}$$

Figure 2.16 Numerical design matrix representation for the matrix shown in figure 2.15.

Two quantitative measures developed by Suh and Rinderle(1982) and Suh(1988) can be used to quantify the degree of functional coupling due to interactions between physical parameters. The two quantities are called Reangularity (R) and Semangularity (S).

Reangularity is a measure of the orthogonality between physical parameters and its metric is given below in equation 2.14 for a 2X2 design matrix.

$$R = \{1 - (A_{11}A_{12} + A_{21}A_{22})^2/[(A_{11}^2 + A_{21}^2)(A_{12}^2 + A_{22}^2)]\}^{1/2} \qquad [2.14]$$

Semangularity characterizes the relationship between a pair of physical parameters and functional requirements. The metric defining this relationship for a 2X2 design matrix is given by equation 2.15.

$$S = [(A_{11})/(A_{11}^2 + A_{21}^2)^{1/2}] \bullet [(A_{22}/(A_{12}^2 + A_{22}^2)^{1/2}] \qquad [2.15]$$

If the design matrix were chosen such that R1 was the physical parameter satisfying the functional requirement Dfs, replacing R_2 in figure 2.16, then the new relations for A_{12} and A_{22} can be derived from equations 2.5 and 2.6 to yield.

$$A_{12} = (\partial Fc/\partial R_1)/Fc = -1/R_1 + (R_2+Rg+Rs)/ (R_2Rs+R_1R_2+RgRs+RgR_1+R_1Rs)$$

$$A_{22} = (\partial Fc/\partial R_1)/D_{fs} = -(Rs+R_1)/(R_2Rs+R_1R_2+RgRs+RgR_1+R_1Rs)$$

The new design matrix is calculated for R_1 replacing R_2 as the physical parameter and represented in figure 2.17.

RC Filter Design Matrix Values with $R_1 = PP_1$

$R_1 = 1060\Omega$, $R_2 = 463\Omega$, $C = 257\mu f$

$$\begin{Bmatrix} F_c \\ D_{fs} \end{Bmatrix} = \begin{bmatrix} -3891 & -0.0000805 \\ 0 & 0.0000805 \end{bmatrix} \mathrm{X} \begin{Bmatrix} C \\ R_1 \end{Bmatrix}$$

Figure 2.17 Numerical design matrix representation for the RC filter in figure 2.6 with C and R_1 as the physical parameters.

The Reangularity and Semangularity measures described in equations 2.14 and 2.15 can be calculated for the designs in figure 2.16 and 2.17 to determine the degree of functional coupling for each design. The results of these calculations are tabulated below in Table 2.9.

Table 2.9
Semangularity and Reangularity Measures of Independence

PPs	S	R
R2, C	0.98	0.954
R1, C	0.707	0.707

From the results given in Table 2.9 it can be concluded that the degree of coupling due to R_1 and C is greater than the coupling due to R_2 and C.

The results obtained here agree with the experimental results obtained when employing the analysis of variance technique reported in Tables 2.7a, 2.7b, 2.7c and 2.7d. The Reangularity and Semangularity measures give a more precise measure of the functional coupling between parameters than the analysis of variance technique.

The disadvantage is that in practice a precise mathematical expression for a design is often not available. And when such an expression exists, it is typically several orders of magnitude more complex than the RC Filter example provided here. However, programs such as MAXIMA that can do symbolic mathematics could be employed to characterize complex designs when necessary.

The analysis of variance and SN measures are generally easier to apply in situations that have many physical parameters and functional relationships that interact in a complex manner, with no known mathematical expression available.

It has been demonstrated here that the metric employed by Suh(1988) and Rinderle(1982) in the application of the independence axiom have derived the same results as the analysis of variance technique employed by Taguchi(1987).

No matter which metric is employed, the fundamental principle of design which says a design should maintain independence of functional requirements is the common principle which guides the designer in the use of such metrics.

2.4 Conclusions

The two design axioms agree in principle with the methods of engineering design described by Taguchi(1987). Their are two obstacles that generally make it difficult to see this point, the language and the metrics.

The language of axiomatics is different than the language employed by Taguchi in describing the design process. Much of Taguchi's analysis technique is an extension of statistical techniques and his language is therefore similar (e.g., factors, interactions). In order to clarify these semantics, Table 2.10 compares the languages with a few examples, in each case showing how the same argument can be made using a different expression.

Table 2.10
Axiomatics and Taguchi Methods, language compared

Axiomatics	Taguchi Methods
Maintain Independence of Functional Requirements.	Minimize interactions between control factors
Minimize Information Content.	Maximize the Signal-to-Noise Ratio
All functional requirements are coupled within some tolerance by physical parameters.	"Interactions always exist to a greater or lesser degree"
If two designs are coupled, the one least coupled is the better one.	"That interaction between control factors is great essentially means that the reliability of experimental results is greatly lowered."
A coupled design can have one unique solution.	"If the interactions are great, no assignment works well except experiments on a certain specific combination."
A coupled design requires more information than an uncoupled one.	"As long as it is not possible to minimize interactions between control factors, it is impossible to render an experiment efficient."

The language of axiomatics is different than that of Taguchi, but the principles are the same. Maintain independence of functional requirements and minimize the information content of the design are the principles.

The second obstacle to understanding the meaning of the axioms is in the metrics used in their application. The metrics as have been employed in this chapter are not the fundamental principles, they are measures employed to gain information about a design in order to apply the principles. Thus, the distinction between an axiom and a method made in chapter I is worth repeating here. That is, methods are context dependent while axioms are true statements.

The information measure described in equation 2.3, the Signal-to-Noise measure in equation 2.8, the sum of the squares variance measure in equation 2.13, the reangularity measure in equation 2.14 and the semangularity metric in equation 2.15 all represent context dependent ways of formalizing information. On the contrary, axioms are, by definition, true statements independent of context.

With this in mind, the question posed for the next chapter is, what would the result be if these axioms were applied to a system of policies and laws instead of an engineering system? The U.S. Patent System has been chosen to test the validity of the axioms in a context well outside the bounds of engineering.

References

Clausing, Don, "Notes on the Signal-To-Noise Ratio", 1988.

Fisher, R.A., Statistical Methods for Research Workers, first edition, Oliver & Boyd, Edinburgh, 1925.

Rinderle, J.R., "Measures of Functional Coupling in Design", Ph.D. Thesis, M.I.T., May 1981.

Scheffé, Henry, The Analysis of Variance, John Wiley & Sons, Inc., New York, 1959.

Suh, Nam P., The Principles of Design, Oxford University Press, 1988.

Taguchi, Genichi, System of Experimental Design, Engineering Methods to Optimize Quality and Minimize Cost, American Supplier Institute, Inc., 1987.

Taguchi, G. and Phadke, M.S., "Quality Engineering Through Design Optimization", IEEE Global Telecommunications Conference, GLOBCOM '84, Atlanta, GA, November 26-29, 1984.

CHAPTER III - THE UNITED STATES PATENT SYSTEM

3.1 Introduction

The purpose of this chapter is to illustrate a method for applying the design axioms to a system of laws and policies. It is proposed here that the principles of design embodied in the independence and information axioms described in chapter II are universal. To test the universality of these principles in contexts other than engineering, the United States Patent System has been chosen.

The Patent System is, like the axioms, concerned with the promotion of innovation in design, which makes it an especially interesting example of policy design. In addition, the U.S. Patent System is currently facing important policy questions in many areas such as the international harmonization of patent laws, protection of new technologies (e.g., living organisms, software, etc.) and the burden of litigation and interference costs on the individual inventor or small business. Thus the Patent System was chosen because of the need to address such questions as well as for its compatibility of interests with axiomatics.

Two schools of thought prevail today concerning the U.S. Patent System. First, there are those who believe the system to be sound in its fundamental constructs, requiring only slight modifications from time to time in order to adjust for changing conditions, new technologies and other needs. This school of thought believes that any necessary changes can be accommodated within the boundaries of the system as it currently exists. This view of the system has been taken by the majority of individuals involved in the design of the system (i.e., Congress, the Courts and the PTO) for more than one hundred years. Some of the more recent changes made within the system include the Patent Law Amendments of 1984 (P.L.98-622), the Semiconductor Chip Protection Act of 1984, the Patent Cooperation Treaty (P.L.99-616), the Drug Price Competition and Patent Term Restoration Act (P.L.98-417), and the Omnibus Intellectual Property Rights Improvement Act of 1987.

Secondly, there are those who suggest that the vast need for legislative alterations within the patent system, especially in the recent past, indicates the existence of a fundamental problem which can only be eradicated through an alteration in the system itself. The United States Congress was given the power to encourage innovation through the protection of intellectual property in Article I, section 8 of the Constitution. Thus, the power to change the system does exists and is in the hands of Congress.

The patent system is in a constant state of redesign. It changes not only with every new piece of legislation, but also with every new interpretation of the law by the courts in their decisions and the Patent and Trademark Office (PTO) in their policies. To represent such a system, taking into account the dynamic nature of its state, a representation of the patent system as a design process is necessary. Only after the system is modeled as a design process with a set of functional requirements or goals can one proceed to apply the design axioms presented in Chapters I and II toward the determination of a proper set of physical and process parameters. Resulting from the application of such axioms to the U.S. Patent System will be a set of recommendations for the future system.

To proceed with this application, a brief historical description of the U.S. Patent System will be provided as an introduction. Then a more detailed introduction to the current set of laws that define the guidelines governing the United States patent system is presented. A design representation for the Patent System will then be elaborated upon with sufficient detail to illustrate the application of the axioms. The independence and information axioms will then be applied toward the evaluation of some recent policy questions in the field of intellectual property. Through the illustration of this design process the utility and context independence of the design axioms will be demonstrated. In addition, this analysis of the patent system will contribute a unique systems perspective to some current policy questions concerning the U.S. Patent System.

3.2 Brief History of the United States Patent System

The United States Patent System was born from Article I, Section 8 of the U.S. Constitution, which reads as follows.

> The Congress shall have the power . . . to promote the progress of science and the useful arts, by securing for limited times to authors and inventors the exclusive rights to their respective writings and discoveries.

This provision resulted from the merging of the ideas of James Madison and Charles Pickney at the Constitutional Convention. This provision was subsequently adopted unanimously without any dissenting views. This one section of the constitution represents the foundation of the U.S. Patent System as it exists today in 1988.

Their are actually two provisions contained in Article I, Section 8. The first secures "for limited times to authors . . . the exclusive rights to their . . . writings" from which our system of copyright laws has evolved. The second provision secures "for limited times to . . . inventors the exclusive rights to their . . . discoveries" from which the current system of patent laws has evolved. The underlying purpose of each provision was to "promote the progress of science and the useful arts".

The first patent law was enacted on April 10, 1790 by the First Congress and was entitled "Acts to promote the progress of the useful Arts". This first law vested the jurisdiction to issue patents to a board led by the Secretary of State Thomas Jefferson along with the Secretary of War and the Attorney General. This board was given the power to issue patents without question or repeal. Under this first Patent Act of 1790, patent litigation or interferences were non existent and the board was given the final word in all cases.

As government proceeded to demand more time of these board members due to their high office, it became necessary to delegate the responsibilities of granting patents. Under the Patent Act of 1793, the granting of patents became a clerical function in which all those who filed an application were granted a patent upon receipt of the proper papers and fees.

This system continued until dissatisfaction in the granting of patents without any examination resulted in a new law enacted on July 4, 1836. This law created a patent office with a commissioner of patents and examiners with the power to refuse patents. Amendments to this law continued through 1870 when the law was completely revised. This law contained the fundamental principles and constructs of the system currently used today in 1988.

The next major revision in patent law occurred in 1952 when a bill for the purpose of codifying the current patent law into Title 35 of the U.S. Code was enacted. This law mainly reorganized and clarified language in the existing law.

For example, section 103 was changed in the 1952 law to include the condition of nonobviousness for patentability. Actually this law had been part of the system for more than 100 years by decision of the courts and was only being formalized in this new Patent Act of 1952.

Since 1952 very few major changes have been made in the patent system. The protection of plant life, semiconductor masks and software embodied in a process as well as adoption of the International Patent Cooperation Treaty are some of the major changes. The majority of changes in the patent system since 1952, and indeed since 1870, can be characterized as changes occurring within the constructs of the existing system. For more than one hundred years the physical make-up (i.e., laws and policies) of the patent system have changed incrementally in response to a variety of social, political and economic pressures while the process or implementation of this system has remained in large part the same.

This phenomenon is interesting in itself when compared to another system such as the production of automobiles. If for example the Ford Motor Company employed the same system for manufacturing automobiles today as Henry Ford did on his first production line, they would be quickly forced out of business due to the higher costs of production relative to their competitors. However, Ford still implements their automobile assembly line technique during production, although the physical make-up of this assembly line has changed dramatically over the years. Ford competes with others to manufacture the best quality and economically affordable automobiles for the customer. The penalty for failure is loss of business and ultimately the destruction of the company. The possibility of this penalty for performing below par gives Ford and others the incentive to continuously improve their design and manufacture. This competitive atmosphere is at the heart of our free enterprise for which the patent system seeks to encourage among inventors.

However, the patent system is different in that it has no competition in the United States. Thus, unlike Ford, their is little incentive for innovations in those designing the patent system because of the lack of competition. It is not meant to say here that there ought to be several patent systems competing with each other in the United States so as to give incentive for better performance, but rather to recognize that this noncompetitive atmosphere can create a tendency to manage rather than innovate. To innovate requires change, which requires more effort than to simply manage the status quo, but which is necessary when designing a dynamic system. This lack of competition may partially explain the lack of change in more than one hundred years to the fundamental constructs of the patent system.

The patent system is a complex system of Laws and Policies that must be organized by purpose in order to develop the system design perspective needed to effect controlled and desired change.

3.3 Laws of the U.S. Patent System

The current set of laws pertaining to patents are organized in Title 35 of the United States Code. These laws represent the efforts of Congress to fulfill their constitutional mandate. Within the bounds of these laws, variation in the patent system takes place each time a judge interprets a law for the context of a specific case or the patent Commissioner makes a policy for the Patent and Trademark Office (PTO) in response to a perceived social, political or economic need.

The interpretations of the courts and PTO policies can change on a daily basis, while new laws legislated by congress change less frequently. The constitution empowers Congress with the responsibility for promoting the progress of science and the useful arts, from which the laws in Title 35 of the U.S. Code follow and following from these laws are judicial interpretations and PTO policies. As part of an introduction to the current patent system, the Acts of Congress organized in Title 35 of the U.S. Code will be briefly outlined.

This Title 35 is divided into four parts with a total of thirty seven chapters embodying U.S. statutes governing patents. The first part has four chapters that govern the organization and activities of the Patent and Trademark Office. The second part contains nine chapters covering the patentability of inventions and the granting of patents. The third part has six chapters of code designed to protect the rights of patent holders. The final part has three chapters that govern the interface of the U.S. Patent law with those Countries that have signed the Patent Cooperation Treaty.

3.3.1 Part I - Patent and Trademark Office

The PTO was established for the purpose of examining patent applications submitted to it against the requirements of patentability described in part II of this title. The policy of the U.S. Patent System and thus the primary goal of the PTO is to encourage invention.

The PTO is part of the Department of Commerce and is head by the Commissioner of Patents. The Commissioner is responsible for the execution of all duties required by law respecting the granting of patents and the registration of trademarks. The Commissioner is also granted the authority to carry out special studies and programs regarding domestic and international patent and trademark law.

The Board of Appeals was established as part of the PTO to review adverse decisions made by examiners upon request of the applicant. As of September 1987, approximately four percent of the 256 thousand pending patent applications were on appeal. During fiscal year 1987 in thirty percent of the 4572 cases decided by the Board of Appeals, the decision of the prior examiner was reversed.

In Chapters 2 and 3 of this title the Commissioner of patents is empowered with the authority to make rules and regulations governing the proceedings and practice before the PTO respectively. The final chapter of this section declares the fees that will be charged by the Commissioner for patent applications, appeals, maintenance and other fees associated with the patenting process. Fees associated with other necessary processing, services or materials may be set by the Commissioner. More than four hundred pages of rules set by the Commissioner of the PTO are published in the code of federal regulations, volume 37.

This first part of the U.S. Code Title 35 defines the PTO through which the day-to-day operations of the U.S. Patent System proceed. In 1987 there were 137 thousand patent applications filed and the PTO issued 88 thousand patents and collected more than $111 million in patent fees.

3.3.2 Part II - Patentability of Inventions and Granting of Patents

The second part of Title 35 U.S. Code defines what constitutes patentable subject matter and the basic procedure to be followed in the examining and granting of patents. Herein are defined the guidelines to be followed by the PTO examiners as well as the courts in issuing patents.

Patentability of Inventions

The first Chapter of this second part is entitles "Patentability of Inventions" and includes five sections that define patentable subject matter. Sections 100 defines the terminology used in this chapter regarding patentable subject matter and section 101 defines patentable subject matter as follows.

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement therefor may obtain a patent therefor, subject to the conditions and requirements of this title. (Title 35, Section 101)

When questions arise as to the interpretation of this statement in the context of a specific invention, the power to make decisions of whether to grant of not grant a patent belongs to the patent examiner, the patent commissioner, the Board of Appeals and the U.S. judicial system. In a decision by the court of Customs and Patent Appeals(1976), it was declared that inventions such as "printed matter, methods of doing business, purely mental steps, naturally occurring phenomena or laws of nature, and mathematical formula and algorithm therefor," are not patentable subject matter. This decision restricts the patenting of software, which has become a billion dollar industry with projections of over $100 billion in revenues for U.S. firms in 1995 made by the Computer and Business Equipment Manufacturers Association (1986).

On March 3, 1981 in a 5-4 vote, the U.S. Supreme Court decided that inventions and discoveries that involve computer programs may be patentable. It was declared that an industrial process does not become unpatentable just because it includes a formula or program. However, the question of how to protect innovations in software not described as part of an industrial process remains as an important issue to be resolved. This question of how to protect innovations in software so to encourage progress of the useful arts is the responsibility of Congress and will be discussed later.

Section 102 of this title declares that an invention must be novel or new in order to be patented. An invention is defined as novel under this section when the following conditions hold true.

1. If the invention was not known, used, patented or described in a printed publication anywhere in the world prior to the established date of invention or for more than one year prior to the date of application for patent in the United States.
2. If the invention is not abandoned.
3. If the invention was not patented by the applicant in a foreign country prior to the date of application for patent in the U.S. or if the applicant has not filed an application in a foreign country more than one year prior to the date of application for patent in the U.S.
4. If the invention was not described on another U.S. or international patent application prior to the established date of invention.

5. If the applicant did indeed invent the subject matter sought to be patented.
6. If the applicants invention was not previously made in the U.S. by another who did not abandon, suppress or conceal it.

Section 102 was first designed to prevent someone from patenting something previously invented elsewhere in the world. Secondly, the one year grace period for filing an application was designed to prevent the removal from the public domain of an invention which has justifiably been thought to be freely available as well as to provide a reasonable time for the inventor to test the invention so to determine whether or not the invention is worth patenting. A grace period before filing is consistent with the innovative design process in which a period of time may be required for the inventor to discover if the invention is in fact useful and worth patenting.

The purpose behind the second part of the novelty requirement is to encourage the prompt use and public availability of the invention or discovery. Abandonment of invention means that the inventor has relinquished the exclusive rights of the invention to the public. The intention of the inventor is the test of abandonment which can be implied by the conduct of the inventor being inconsistent with the intention to claim a patent.

Another purpose of the novelty requirement is to prevent an inventor from taking advantage of patent rights for a substantially greater time period than the allotted seventeen years. It is also designed to award patents to the inventor who can establish the earliest date of invention as opposed to awarding it to the first to file a patent application as is the practice of the European Patent Organization and other countries patent authorities as well. Determining the first-to-invent can be difficult and it is typically done through presenting evidence such as dated notes, publications and/or individual testimony to support claims of invention. The issue of going to a first-to-file system in the U.S., thus redesigning section 103 of this title, is a major consideration for the U.S. patent system in the current effort to harmonize the U.S. system with those in the rest of the world and is discussed later.

In summarizing the purpose of section 102 of this title on novelty, the two major requirements of this section can be described as follows.

Major Functional Requirements of Section 102

1. To protect the rights of the individual(s) who was the first inventor by awarding a patent to and only to said inventor.
2. To protect the rights of the public by preventing the exploitation of inventions for more than the allotted 17 year period by the inventor.

The requirements for patentability have one more major provision, which is described in section 103 of this title as the nonobvious subject matter condition. Under this section a patent may not be obtained if,

> the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains.

The purpose of this section is to prevent the award of a patent for an invention that is an obvious extension of prior knowledge. If an invention fulfilling both the utility and novelty requirements for a patent was truly obvious to any individual having ordinary skill in the art of the subject matter to which the invention pertains, then there would likely be several other patent applications or interferences filed for that same invention. If there are one or more prior inventors for the alleged obvious invention, the rights of those inventors are protected under the novelty requirement in section 102 of this title. Hence, the purpose of section 103 would appear to be redundant with section 102 in that an invention that is obvious is unlikely to also be novel.

This requirement for obviousness is extremely difficult for individuals, skilled in the art or not, to judge. When questions of obviousness are posed to designated experts in particular technologies, the advantage of hindsight can easily haze an individuals judgement. The purpose of the patent system as a whole must be examined in this situation, since the intent of the system is not simply to award patents to individuals with ideas, but rather to innovators who turn their ideas into useful products, processes and other useful subject matter. Thus, although nonobviousness may theoretically be a valid requirement for patentability, the inability to make such a judgement could potentially hinder the system in fulfilling its primary purpose, the encouragement of innovation.

New areas of biotechnology, genetic engineering, information science and other technologies are currently challenging the ability to judge obviousness. For example, if drug A exists and was developed for and is used to treat disease B, and then another discovers that drug A can be used to treat disease C, is the discoverer of the new application entitled to a patent? When a person of ordinary skill in the art is presented with the necessary information, that individual may declare the new application to be obvious.

The innovation or design process is itself simply a process of organizing information in such a way that solutions to problems, inventions or designs become obvious to the inventor. The ability to organize information and not simply the ability to solve a problem with all the facts on the table is an important quality of an innovator that must be recognized and rewarded. The obviousness requirement of section 103 judges based on an experts opinion with the facts already before him. If it were not obvious to the expert at that point, he would not be an expert, and to ask an individual to step back in time assuming no knowledge of the information may be impossible given the constraints of human nature.

In the patent infringement case of Bird Provision Co. v. Owens Country Sausage (568 F.2d 369), in February of 1978, the U.S. Court of Appeals affirmed an earlier decision that the patent held by the Bird Provision Co. was invalid due to violations of the novelty and obviousness requirements of patentability. It was shown in this case that the patented process for packaging sausages, although of great utility, had been used by several other packaging companies prior to its use by the Bird Provision Company. It was also subsequently declared to be obvious. In this case the requirement for novelty and for nonobviousness were both violated. Are these two requirements independent of each other? If for example the patent was determined to be novel and was not used prior to the date of patent, would it nevertheless be declared obvious? It can be concluded that an invention of great utility with no prior use is consequently not obvious.

In another case involving the application for a patent by Leslie(1977), the subjectivity of the judge in the decision making process was pointed out in the notes of this case concerning obviousness. The notes stated that the,

> issue of obviousness of a given design is ultimately resolved subjectively according to the visual perception of the judge who finds himself cast in the role of the ordinary observer.

The final section in this title concerned with the definition of patentable subject matter is a clause which affords equal privileges to inventions made in foreign countries by individuals in service of the U.S. Government. As the cooperation of international patent authorities progresses, this statute may become redundant and consequently unnecessary.

The statutes concerning the patentability of inventions, namely sections 100, 101, 102 and 103 of this title will be tne main focus of this Chapter and have thus been elaborated upon in this introduction. Some major issues of immediate concern to the U.S. policy concerning the protection of intellectual property rights in fulfillment of Article I, Section 8 of the U.S. Constitution have been identified here. Issues such as the patentability of software and other new technologies, the awarding of patents to the first inventor or first-to-file, and the recognition of the innovation or design process within the patent statutes will be subsequently analyzed. Recommendations for the U.S. patent system based on the design axioms will follow this analysis.

Granting of Patents

The next four sections of Title 35 define the details of patent application content required for processing, the requirement for examination of proper applications to be carried through by the Commissioner of the PTO, the right of an applicant or party to an interference to appeal the PTO decision to the Court of Appeals for the Federal Circuit and the requirements for issuing of patents after the technical validity of a patent is determined by satisfying sections 101, 102, and 103 of this title. The reason that PTO decisions on patents are appealed to the courts is because of the wide margin of uncertainty associated with the meaning of the law in the context of a particular invention. The costs of litigation in these situations is high for both the individual inventors and the public. Thus a major challenge for designing the patent system is to reduce the uncertainty associated with the meaning of the statutes so as to minimize the total amount of time and resources spent on litigation.

In section 161 of this title organic plants are declared as patentable subject matter. On June 16, 1980 the Supreme Court decided in a 5-4 vote that living organisms are patentable under existing U.S. law. This decision allowed Mr. Chakrabarty a patent for a bacterium produced in the laboratory and capable of breaking down crude oil. More recently genetically engineered plants and animals have also been patented. New questions involving the ethics of patenting living organisms and animals as well as questions of protection or enforcement of such patents have arisen.

Section 171 of this title authorizes the granting of patents for any "new, original and ornamental design for an article of manufacture." The purpose of this section is to encourage the decorative arts and thus the requirement of utility is eased for these cases.

Section 181 of this title provides for the national security interest of the U.S. by authorizing the withholding of patent information from the public if determined that doing so is in the national interest. This is designed to prevent the patent system from interfering with the broader requirement of the government to provide for the safety of all citizens.

The final group of sections in this title, beginning with section 200, defines the patent rights for inventions made with federal assistance. The purpose of these sections are to encourage and give incentives for individuals to utilize ideas in the form of new inventions which have come about through federally funded projects.

3.3.3 Summary

This completes a brief description of the first two parts of the U.S. patent system as described in Title 35 of the U.S. Code. The description of the patent system is divided into four functional parts: (1) to define the duties and physical makeup of a Patent and Trademark Office, (2) to define what constitutes patentable subject matter, (3) to define the manner in which those rights will be protected, and (4) to define an interface between the U.S. Patent System and other foreign patent systems. The issues focused on in the remainder of this thesis are concerned mainly with the definition of patentable subject matter and the protection of patent rights. Therefore, the brief description of the remaining two parts of the patent system can be found in appendix 3A at the end of this chapter.

The previous outline of the U.S. patent system is an attempt to describe not simply the law as it reads, but also the purpose or intent of the particular statutes. Certain parts of the system were elaborated on more than others for the purpose of highlighting issues in the protection of intellectual property that will be addressed further in this thesis.

To address the issues concerning the patent system, the fundamental approach to design as described in the first two chapters of this thesis will be used. In order to utilize the principles of design in this approach it is first necessary to develop a design representation for the patent system.

3.4 Representing The Patent System as a Design Process

3.4.1 Why Represent the Patent System as a Design?

The purpose of the design representation illustrated here is to present a perspective of the patent system which includes not only its structure (i.e., laws and policies), but also the reason for the structure. An understanding and thorough awareness of the functional requirements (FR) for the patent system is needed to design laws and policies to effectively fulfill such requirements.

It is the common experience of system designers, that FR's change over time. This requires that the physical parameters (PP's) must also change to accommodate the proper functioning of a system. However, if a designer changes a PP in one part of a large system to accommodate a new FR, and the new PP is coupled with(i.e., interdependent on) another FR somewhere else in the system, then the original change will cause a disruption in another part of the system. This disruption will initiate cause for a modification in that part of the system adversely effected by the initial modification. This new modification will cause another disruption, then another, and still another in a recursive loop. In fact one change could provoke two or three, which could lead to an avalanche of small perturbations in the system.

This process of change may be perceived as normal, since in any large system, modification is a natural process in response to a changing environment. However, in order to responsibly change a large system, it is not satisfactory to only consider a small part of the system, but the entire system must be considered as a whole.

This "process of change" is in fact the design process, and in order to achieve a representation for such a process, it is not adequate to model the structure of a system alone, but in addition to the structure (i.e., PP's), a representation of the reason for the structure (i.e., FR's) is a necessity for those who seek to responsibly alter a system.

It is common to see organizational systems modeled in one dimension, describing the structure, but not the reason for the structure. However, systems that continually "change", in response to a changing environment, require representation as a "design" process. By representing the U.S. Patent System as a complex design task, important questions and issues will become evident, and recommendations for long term solutions will be revealed.

3.4.2 Who Designs the Patent System?

Their are many actors who contribute to designing the patent system. The United States Congress holds the primary responsibility for the patent system design through the legislation of laws as is their constitutional mandate. Congress has delegated responsibilities for parts of the system to both the executive and judicial branches of government. A depiction of the patent system designers in the U.S. government is shown in figure 3.1.

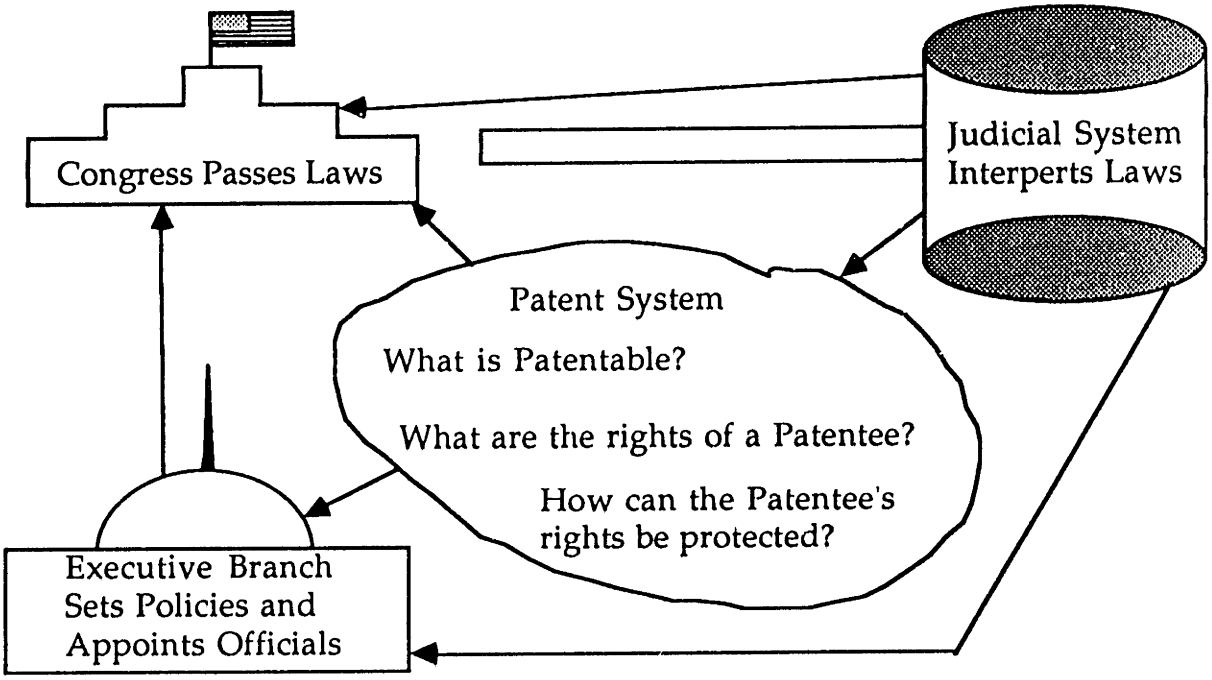


Figure 3.1 The principle designers of the U.S. Patent System are illustrated here, showing their interrelationship and basic tasks in the design process.

Congress retains the primary power to change any and all fundamental aspects of the patent system that are within its constitutional mandate. Congress manifests this power through the legislation of laws that define the physical embodiment of the patent system. Changes in legislation can take months, or more typically years, and thus the response time of congress in addressing patent system problems resulting from a rapidly changing environment can be slow. To deal with the patent system design and maintenance with a faster response time, Congress has delegated authority to both the executive and judicial branches of the U.S. Government.

A Patent and Trademark Office (PTO) was established as part of the executive branch to deal with the day-to-day maintenance of the system as well as the continual redesign of the system. The Commissioner of the PTO has the authority to set policies that govern the issuing and protection of patents within the constraints of the Congressional legislation. Based on the guidelines given them by congress, the PTO must distinguish between inventions and non-inventions in order to grant patents. The Commissioner of the PTO is given a great deal of flexibility to design a system in which to implement the congressional guidelines. Each time a new policy, rule or regulation for the PTO is declared, the patent system is changed. The response time for the Commissioner in implementing changes in the patent system design is typically much faster than Congressional response time, although the magnitude of the change is more restricted.

Congress has also given the judicial system a role in the design of the patent system by giving it the authority to settle disputes arising from uncertainty in the meaning of the its own legislation. The judicial system is also given the power to inflict punishment on individuals or groups that disobey the patent laws designed by congress or the policies and regulations of the PTO. Whenever a judge makes a decision on a patent case by interpreting a Congressional Law or PTO policy in a particular context, a precedent is set and the patent system is consequently altered. Thus, the judicial system is responsible for designing the patent system each time an interpretation of the congressional intent is made through deciding a patent case in court. The response time for the judicial system in making interpretations is typically faster than congressional response time, but it can be delayed a great deal due to constraints on the judicial system itself.

The three major participants in designing the patent system are presented here as being the three branches of government. Each participant plays a different role in the design process. The dynamics of this process can be likened to the traditional three players in an engineering design process as the design group, manufacturing group and the customer service group. Congress designs the patent system and hands off their design to the manufacturing group (i.e., PTO) to be implemented. The PTO produces a product, which is more in the form of a service, for which their customers are inventors. When the design fails in the field, the judicial system acts as a customer service representative who must fix the system or simply demonstrate to the customer how to use the system properly.

Typically, in engineering design, after the customer service engineer has put enough patches on the current design in the field, a redesign of the system is initiated in order to incorporate changes made in the field into a new and improved product. The judicial system preforms an analogous function for the patent system in patching it in the field, after which if enough judicial decisions are made, that result in a standard precedent for the interpretation of the congressional intent, the congress may incorporate such precedents in a new piece of legislation. In this way, designing the patent system is very similar to designing a product.

A product designer strives to design so to facilitate the manufacturing process. The hand-off of designs from the design group to the manufacturing group is often a major source of problems for corporations involved in product design. The cost of addressing problems in the design stage is considerably less expensive than during the manufacturing stage. The cost of solving product problems after the manufactured product has been sold to customers is even more expensive and in some cases can result in the demise of the company. Subsequently, it is much more cost effective to concentrate on improving the design process than it is to expand the customer service organization. Many companies are following this remedy and devoting significantly more effort into designing products for manufacture so to eliminate the need for customer service.

This analogy to product design is similar for the patent system. The cost of litigating a patent dispute in the courts is more costly than settling the patent application through the Board of Appeals, the reexamination process or arbitration. The cost of solving a problem during the legislation of a law is less expensive than the cost of solving the problem during the implementation process. Based on this analogy it would seem more cost effective to focus more resources on the problem solving at the legislative stage in the design process so as to simplify the implementation and minimize the need for the judicial system (i.e., customer service).

Another major player in the patent system design, perhaps most important, is the customer. Just as failure to recognize the needs of the customer in a product design can result in poor products, the failure to recognize the needs of the customer in the patent system(i.e., the innovator) will result in a poor patent system design. The patent system must be designed to fulfil the needs of innovators. If the patent system is to be designed to encourage individuals to innovate and invent new products and processes, input from such individuals must be solicited early in the design stage (i.e., during the legislation of laws).

The designers of the patent system have been categorized into four groups with varying spheres of authority and influence. The four groups consist of: (1) the United States Congress, (2) the Patent and Trademark Office, (3) the U.S. Judicial System, and (4) inventors.

3.4.3 A Spatial Design Representation for the U.S. Patent System

Design is a reflective process in which one must first determine the functional requirements of that which is to be designed from an often unstructured environment. Then a designer must develop physical parameters that fulfill the objectives of the functional requirements. Finally, process variables must be designed that will realize (i.e., manufacture or implement) the designed physical parameters. The process involves the continual transition between the functional, physical, and process spaces, as the designer must constantly reflect on the structure of the design, as well as the reason for the structure and the process of implementation or manufacture. The design representation given in figure 1.1 is presented here again in figure 3.2 to help illustrate a spatial representation of design in the context of the patent system.

Patent System Design Space Representation

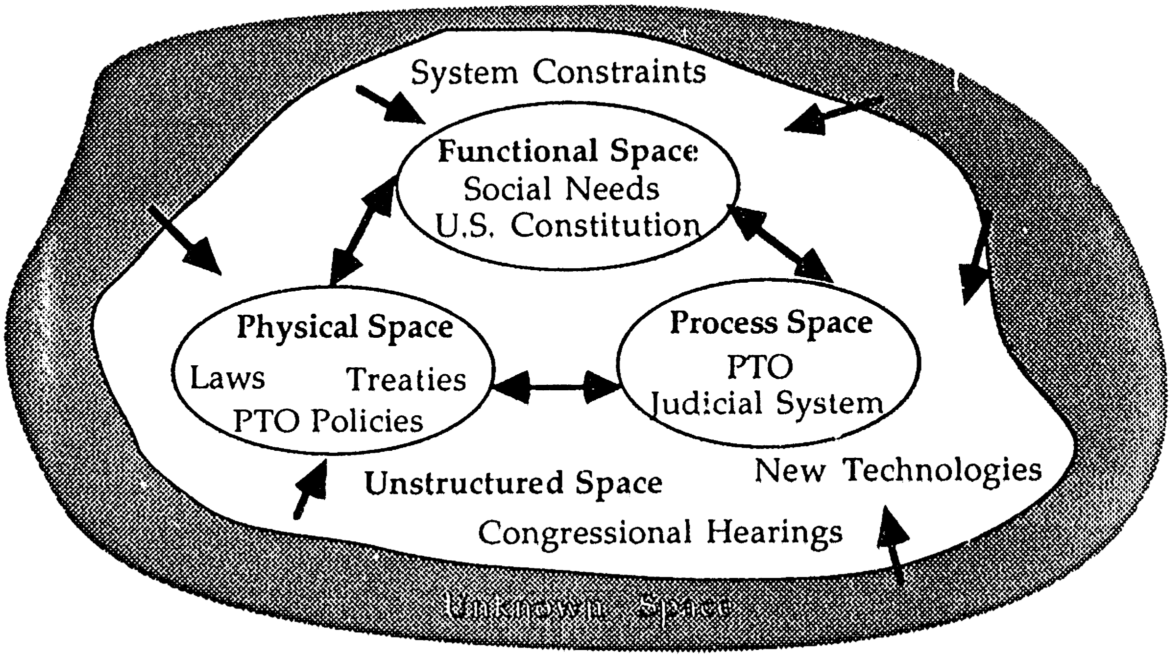


Figure 3.2 A spatial representation of the patent system design process illustrating the causal interaction between the functional, physical, process, unstructured and unknown design spaces.

The unknown space is depicted in figure 3.2 as surrounding the process, physical, functional and unstructured spaces. It represents that which we do not fully understand. As new knowledge is attained, new information becomes available to the designer in the unstructured space from which new functional requirements are often formed. For example, in the context of the U.S. Patent system of 1952, software was a new technology lying in the unstructured space. The protection of software rights was something for which we had little knowledge or even conception, and therefore functional requirements and subsequently laws for the protection of software rights had yet to be conceived or developed. The protection of software has evolved as a major problem for the customer(i.e., the inventor) in 1988. Customer service (i.e., the judicial system) has been brought in to deal with this design problem and many software patent cases have been litigated in the courts. Other new technologies fall into the same category as software such as living organisms and drugs. To design a patent system that can respond to such new technologies, and understanding of the inventive process in such new areas is needed to distinguish between an invention and a non-invention. The current legislation leave "grey areas" in the law for which the judicial system must decide if such new technologies were meant by congress to be or not to be patentable. An illustration of this is shown in figure 3.3.

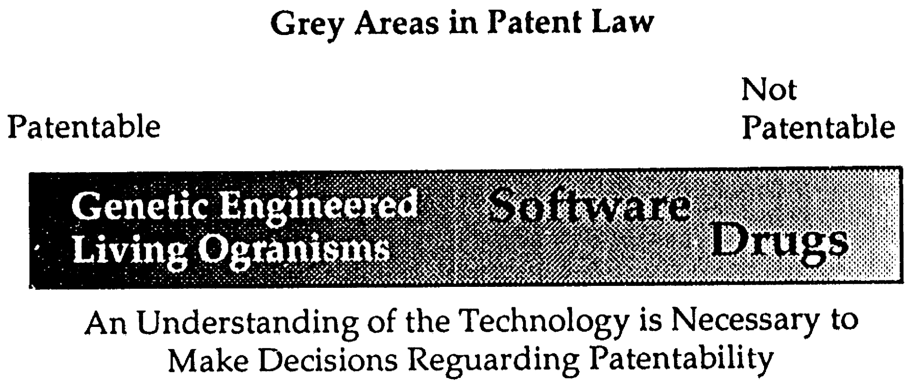


Figure 3.3. A simple illustration of the Black & White nature of the legal system in deciding on the patentability of new technologies.

For the patent system, the process space is the judicial system and executive agencies that define the process through their interpretations of the law (i.e., physical parameters) in both legal decisions and policy rules and regulations. Process variables describe the means by which the physical parameters can be realized or implemented. The rules and regulations for examining and processing patent applications made by the PTO make up the patent system process space.

In the context of the patent system, the functional space would represent the reason for the existence of a patent system beginning with the most fundamental functional requirement given by the Constitution. After the most fundamental requirement is established, a physical parameter must be designed to achieve this requirement.

The most fundamental FR for the patent system is to "promote the progress of science and the useful arts" and the physical parameter chosen to achieve this FR is to secure for inventors the exclusive rights to their inventions.

Functional requirements for the patent system are the congressional intent for the statutes. The statutes themselves form both physical parameters and process variables. The Congressional intent or purpose sometimes seems obvious from observing the statute and other times it seems unclear. Often the purpose or intent behind a particular statute can be found in records of congressional hearings and legislative histories. Section 3.3 of this chapter describes the purpose or intent behind the various statutes of the patent system based on such records of their legislation.

The statutes that govern the patentability of inventions described in section 3.3.2 are considered physical parameters in this representation. The statutes that govern the operations of the Patent and Trademark Office in section 3.3.1 are in large part considered process variables. After the designer specifies a physical parameter(i.e., Law) to fulfill a particular functional requirement (i.e., Congressional intent), a process variable (i.e., PTO policy) must be designed to implement the specified law.

During this design process, one must be mindful of changes in the unstructured space which could change the functional requirements defined earlier. When this happens, the design must be evaluated for consistency between the functional and physical spaces. As one proceeds down the hierarchy of the design process, the prior functional requirements become constraints on any lower level parameters.

3.4.4 Constraints on the Patent System

The patent system is part of a larger system with its own set of functional requirements. Figure 3.4 sketches how the fundamental functional requirement for the patent system fits within the larger framework of the U.S. system of government. It is important to view the patent system from this perspective, because the higher level FR's and PP's of the government system can act as constraints on the patent system design.

The Patent System is Part of a Larger System

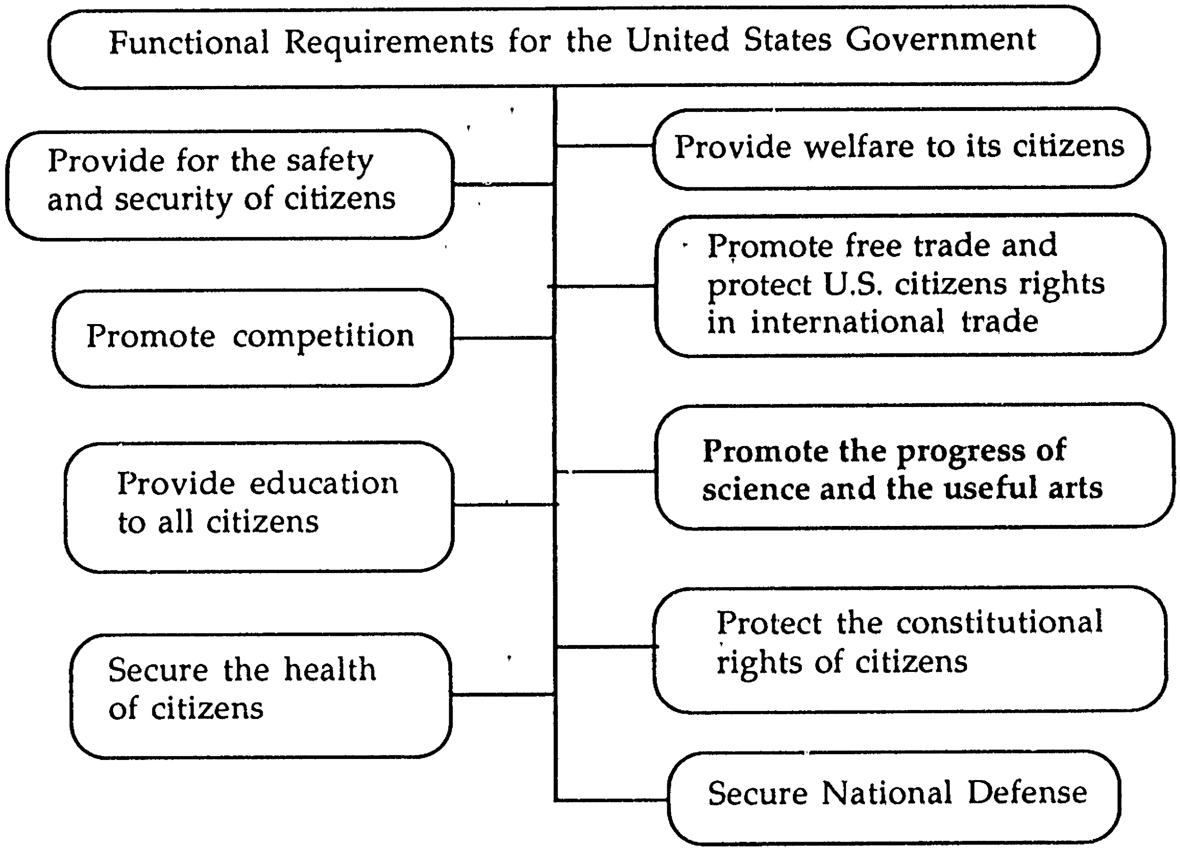


Figure 3.4 The government system in which the patent system functions places constraints on its design. The above diagram is not meant to be a comprehensive treatment of the functional requirements of the U.S. government, but rather to recognize the broader framework in which the patent system resides.

The patent system has been represented within the system of the U.S. government as it currently operates. The upper level functional requirements (e.g., education, security,... etc.) become constraints on the patent system as one proceeds down to a more detailed design representation. For example, the patent law must incorporate national security constraints into it's design by keeping technological innovations that affect national security secret, while still fulfilling the main functional requirement of promoting the advancement of science and the useful arts.

Since constraints are derived from prior functional requirements of the larger system, if an upper level FR shifts due to a change in the unstructured space (i.e., environment or circumstances), a subsequent change in the constraints effecting the lower level hierarchy of FR's, PP's, and PV's occurs. Thus, an awareness of how the government system changes, in which the patent system is only a small part, is important for individuals involved in changing or designing the patent system.

This rippling effect of changes in constraints can be slow in reaching the lower levels of the detailed design. When such delays occur, they are manifested as absent or outdated laws, policies and/or rules. For this reason, it is important to clearly state the FR's, PP's, and PV's of as much of the system as possible, so that those responsible for the design at the more detailed levels will have a clear understanding of the constraints placed upon their task.

3.4.5 Hierarchy of Patent System Functional Requirements

Figure 3.5 shows a hierarchical organization for the functional requirements of the U.S. patent system. This representation has been derived from the statutes in Title 35 of the U.S. Code governing patents. There is a physical parameter and process variable that corresponds to each functional requirement in the form of statutes from the U.S. Code or PTO policies. These statutes have been discused in section 3.3 of this chapter. A more thorough treatment of each individual statute can be found in Title 35 of the U.S. Code, but because of the large volume of code incorporated in this title, the statutes will be treated in detail only when they relate to a specific policy question.

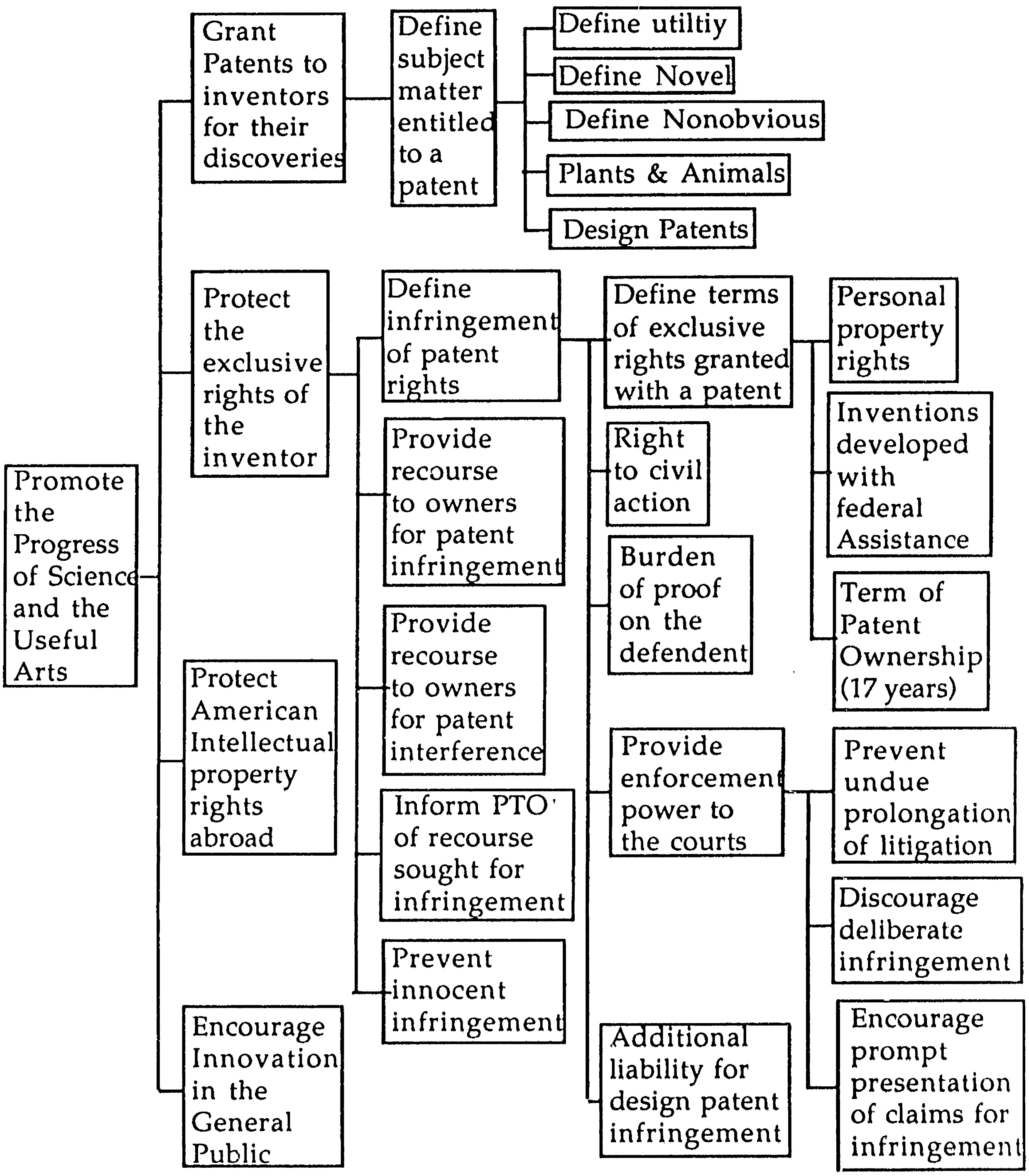

Patent System Functional Tree

Figure 3.5 A hierarchical representation of the functional requirements for the U.S. Patent System.

3.4.6 Summary of the Patent System Design Representation

The patent system has been organized as a system of functional requirements, physical parameters and process variables which exists within the design of the larger system of the U.S. Government. This system of government places constraints on the patent system that must be understood by the individuals who design the patent system. All three branches of the U.S. government play a role in designing the patent system, with the legislative branch holding the primary responsibility for the design, the executive branch (i.e., PTO) responsible for the implementation and the judicial branch taking care of customer service when the design fails to function smoothly.

A hierarchical representation of the patent system functional requirements was developed as shown in figure 3.5. This representation will be used to facilitate the application of the design axioms to designing the patent system in the following sections.

3.5 Application of the Independence Axiom

Based on the application of the independence axiom in Chapter II, a design with little or no interaction between its functional requirements in the physical space is better than one that has many interactions. When a design in which a functional requirement interacts with the physical parameters of other functional requirements, it is said to be coupled to those functional requirements through the physical parameters. The patent system must be observed from two perspectives in order to locate sources of coupling and suggest remedies. First the patent system must be studied as part of the larger system of government and secondly, as a system design itself.

3.5.1 The Patent System as part of the Government System

The patent system represents one functional requirement in the larger system design of the U.S. Government as was suggested in figure 3.4. A design matrix can be developed for this system just as it was in figure 2.2. The design matrix representation for the U.S. government shown in figure 3.6 is not meant to be a comprehensive matrix, but rather one that illustrates the key areas of coupling with the patent system.

A Design Matrix For The U.S. Government

$$\begin{Bmatrix} \text{Provide for the security of citizens} \\ \text{Promote competition} \\ \text{Promote free trade} \\ \text{Provide education to all citizens} \\ \text{Secure the health of citizens} \\ \textbf{Promote the progress of science and the useful arts} \end{Bmatrix} = \begin{bmatrix} \checkmark+ & & & & & \\ & \checkmark+ & & & & \\ & & \checkmark+ & & & \\ & & & \checkmark+ & & \\ & & & & \checkmark+ & \\ \checkmark & \checkmark & \checkmark & \checkmark- & \checkmark & \checkmark+ \end{bmatrix} \times \begin{Bmatrix} \text{Defense Department} \\ \text{Anti-Trust Laws} \\ \text{U..S. Trade Representative} \\ \text{Department of Education} \\ \text{Department of Health and Human Services} \\ \text{Patent System} \end{Bmatrix}$$

Figure 3.6 A design matrix showing a qualitative relationship between the patent system and other functions of the U.S. Government. A check-plus indicates a strong relationship, a check-minus indicates a weak relationship and a check indicates a moderate relationship.

The relationship between functional requirements and physical parameters must be discussed more qualitatively for the patent system than was done for the filter example in chapter II due to the absence of a mathematical expression to describe the patent system design matrix.

Providing Security

The mechanism of the patent system is coupled to the governments goal of providing for the security of its citizens. Information that may jeopardize the security of the United States against its enemies must be withheld from being patented and thus prevented from entering the public domain. Section 181 of title 35 in the U.S. Code guards against the release of such information. This provision has been utilized extensively in the area of nuclear technology.

The implementation of this provision is carried out through the PTO in coordination with the Department of Defence, the Atomic Energy Commission and all other defence agencies of the government. The Commissioner of the PTO is responsible for recognizing suspect patent applications and subsequently submitting them to the above stated agencies for review. The application of this legislation serves to minimize the coupling between the patent system and the system of national defence.

Promoting Competition

The granting of patents to inventors secures, for a limited time, the exclusive rights to their inventions. Section 261 and 262 of title 35, U.S. Code, define the inventors ownership of the patent to be as personal property, allowing the inventor to license, sell or otherwise dispose of the patent at any time during the allotted 17 year patent life. Furthermore, section 154 of this title provides patent owners with the right to exclude others from making, using or selling their inventions. The application of these statutes have caused some contention with the U.S. Antitrust laws that are designed to promote competition.

Section 3 of the Clayton Act contained in title 15 of the U.S. Code declares it unlawful to lease or sell,

> goods, wares, merchandise, machinery, supplies or other commodities, whether patented or unpatented . . . on condition . . . that the lessee or purchaser thereof shall not use or deal in goods . . . of a competitor or competitors of the lessor or seller, where the effect of . . . such conditions . . . may be to substantially lessen competition or tend to create a monopoly in any line of commerce.

The first adverse affect due to the coupling between the antitrust laws and the patent system functional requirement was noticed by the customer service engineers for the patent system (i.e., the courts) in 1912. In Henry v. A.B. Dick Co. (224 U.S. 1, 1912), the court upheld the patent owners practice of requiring the lessee to purchase the patent owners supplies along with the patented mimeograph machine. This decision enhanced the coupling between the two functional requirements by giving the patent owner a monopoly on the supply business as well as the patented machine.

This was not the intention of the patent system and in 1917 the courts saw the fault in the previous decision. In Motion Pictures Patent Co. v. Universal Film Manufacturing Co. (243 U.S. 502, 1917), the owner of a film feeder patent for motion pictures that was vastly superior to others at the time, sought to license said feeder on the condition that the lessee use films from only those individuals approved by the patent owner. In this landmark decision of the court, the patent owners right to impose said conditions on the sale of the patent was determined to be beyond the scope of his patent rights. The doctrine of patent misuse was born by this decision.

The testimony of Susman and Gerlach(1987) before the Subcommittee on Patents, Copyrights and Trademarks examined the history of the doctrine of patent misuse and showed that it had expanded in scope beyond even the existing constraints of the antitrust laws. In Morton Salt Co. v. G.S. Suppiger Co. (314 U.S. 488, 1942), the Supreme Court found that Morton Salt Company's conditional license of their patented salt machine requiring the purchase of their salt tablets constituted a "misuse" of the patent. Furthermore, the court found it unnecessary to show violation of antitrust laws in order to establish patent misuse.

In that 1942 decision the court found misuse based on the showing of some anti-competitive effect and not specifically based on a direct violation of the antitrust statutes. Based on this doctrine of misuse, the courts have designed additional antitrust or anti-competitive constraints into the system. By applying a different standard of anti-competitive behavior to patents as is applied to other cases, the patent system is coupled to the functional requirement of promoting competition. The coupling can be minimized by reforming the current practice of the patent misuse doctrine to determine misuse only for patent owners that specifically violate the antitrust laws.

The antitrust laws are based on the requirement to promote competition and the patent system is concerned with the requirement of promoting innovation. Thus, the patent system should not be concerned with the redefining "anti-competitive" actions, but rather in applying the already defined antitrust constraints to patents.

As part of the Reagan administration domestic intellectual property initiatives, the "Omnibus Intellectual Property Rights Improvement Act of

1987" was outlined in the President's State of the Union Message. Provided in this act is a reform of the patent misuse doctrine to declare that no misuse exists unless the practice violates antitrust statutes. Based on this analysis, the coupling between the requirements of the patent system and the requirement of promoting competition would be minimized by adopting this provision.

Promoting Free Trade

> Worldwide abuses in the intellectual property area may be as blatant as open piracy or as subtle as procedural red tape. In any form, they all have distorting effects on international trade.
>
> Quigg (1987)

The promotion of free trade between the U.S. and other foreign countries is enacted through the U.S. statutes governing trade and through the policies and rules of the U.S. Trade Representative. If these laws and policies inhibit the patent systems ability to promote the progress of science and the useful arts, the two functional requirements are said to be coupled as indicated by the check in the design matrix of figure 3.6.

In the most recent redesign of the patent system defined by the Omnibus Trade and Competitiveness Act of 1988, a source of coupling between these two requirements concerning process patents was resolved. Title IX, subtitle A of this act entitled "Process Patents" amends section 154 of the U.S. Code title 35 by adding to the rights of process patent holders,

> the right to exclude others from using or selling throughout the United States, or importing into the United States, products made by that process.

Prior to this statute, a process patent could be violated outside the U.S. and the product of the process could be freely imported. An additional process variable was also needed to implement this new physical parameter. As a result, a new section 295 was amended to title 35, U.S. Code, declaring that a product be presumed to have been made with the patented process,

> and the burden of establishing that the product was not made by the process shall be on the party asserting that it was not so made.

If the burden of proof were reversed, the application of this new process patent statute would be very difficult, costly and time consuming. It is much easier for a manufacturer of a product to demonstrate that the patented process was not used in its manufacture, than it would be for anyone else. Thus, the new laws governing process patents are successful at decoupling part of the patent system from the requirement of promoting free trade.

Although the process patent protection helps uncouple the patent system requirement of "securing for limited time to inventors the exclusive rights to their discoveries" from the promotion of free trade, it has by no means relieved all sources of contention between these requirements. Hill(1985) of the International Trade Administration reported the findings of the International Trade Commission that over $8 billion and 130,000 American jobs were lost in 1982 due to the counterfeiting of intellectual property. The International Anti-Counterfeiting Coalition and the U.S. Customs Service placed estimates closer to $20 billion.

The protection of patents from counterfeiting can not be designed and implemented within the confines of the physical and process spaces of the patent system design thus far discussed. In order to design a solution to this problem, the cooperation of the world community is needed, since the majority of counterfeiting activities take place in countries where the U.S. has no jurisdiction and only little influence. An international law must be designed in order for the U.S. patent system to secure the rights of inventors internationally.

Thus far there has been some progress made in this direction, but their has yet to be a proposed law or treaty to deal with this issue. The World Intellectual Property Organization(WIPO), which is a branch of the United Nations headquartered in Geneva, convened a panel of experts for the first time in May of 1986 and then again in 1987 and 1988 for the purpose of developing a model law to protect against counterfeiting. In addition to WIPO, at the next meeting of the General Agreements on Tariffs and Trade(GATT) Ministers, negotiations toward the development of a multinational framework of rules dealing with international trade in counterfeit goods will be included.

The promotion of free trade is thus coupled with the protection of patent rights. Such coupling can only be expected to increase as international commerce increases and as the U.S. industry begins to rely more on foreign markets. As this happens, the need for a comprehensive set of international laws governing intellectual property rights will emerge as a top priority for trade negotiators, especially in the industrialized countries. As this process unfolds, the patent system designers as depicted earlier in figure 3.1 will be amended by the addition of the international community through both the WIPO and GATT.

Providing Education

The patent system is not coupled to the functional requirement of the government to provide education to its citizens. However, the patent system does have some complementary goals related to educating youth. In 1987 the PTO began PROJECT XL as a formal program dedicated to developing school curricula to teach analytical thinking and problem solving to school children in the United States. The goal of the program is to encourage invention in the next generation through the early development of related skills.

Although this program certainly has merit, the constitutional mandate of the patent system is to promote innovation by securing for inventors the exclusive rights to their inventions. It does not dictate the development of educational programs to encourage innovation in youth and it does so for a reason. The reason is that in the design of the U.S. government as depicted in figure 3.6, the responsibility for educational programs was given to the Department of Education. Therefore, based on this systems perspective to the government design, the PROJECT XL program should be shifted under the jurisdiction of the Department of Education in order to avoid redundancy in the system.

Securing the health of citizens

The patenting of inventions that could potentially harm the health of citizens such as food, drugs and medical devices must be checked for faults before making them available to the public. The current process of protection from such inventions is implemented by the Food and Drub Administration (FDA) of the federal government. The time needed to evaluate such inventions for safety by the FDA caused a delay in the patent system and coupled the requirement for granting patents.

In September of 1984 the Congress responded to this coupling in the system by passing the "Drug Price Competition and Patent Term Restoration

Act" (P.L. 98-417). Prior to this law, the duration of a patent life included the time needed to test and guarantee the safety of the product. The regulation administered by the FDA to secure the safety of the product can take as long as three years or more for some inventions. Drugs can be expensive to test and develop, thus in order to promote innovation in such technologies while protecting the health of the citizens, this act was passed.

In addition to these specific technologies that pose a large public threat such as food, drugs and medical devices, the potential for harm to come to individuals who use inventions poorly designed and not thoroughly tested exist in all technologies. The question of liability for a faulty invention comes up in such cases.

3.5.2 Coupling within the patent system

The patent system is not only part of a larger system, but also a design in itself. By examining the patent system design for coupling between its functional requirements in both the physical and process spaces, problem areas in the design can be recognized. To aid in this examination, two design matrices have been constructed to illustrate areas of interaction between the functional requirements of the patent system.

U.S. Patent System Design Matrices

$$\text{Functional Requirements} \xrightarrow{\qquad\qquad} \text{Physical Parameters}$$

$$\begin{Bmatrix}\text{Grant Patents to inventors for their discoveries}\\ \text{Protect the exclusive rights of the patent}\\ \text{Encourage innovation in the general public}\end{Bmatrix} = \begin{bmatrix} \checkmark & \text{O} & \text{O}\\ \text{O} & \checkmark & \text{O}\\ \text{O} & \text{O} & \checkmark \end{bmatrix} \begin{Bmatrix}\text{Patentability requirement statutes}\\ \text{Patent infringement statutes}\\ \text{Patent publications and special programs}\end{Bmatrix}$$

$$\text{Physical Parameters} \xrightarrow{\qquad\qquad} \text{Process Variables}$$

$$\begin{Bmatrix}\text{Patentability requirement statutes}\\ \text{Patent infringement statutes}\\ \text{Patent publications and special programs}\end{Bmatrix} = \begin{bmatrix} \checkmark & \checkmark & \text{O}\\ \checkmark & \checkmark & \text{O}\\ \text{O} & \text{O} & \checkmark \end{bmatrix} \begin{Bmatrix}\text{Patent Examiner Board of Appeals Quality Review}\\ \text{Appeal to PTO Civil Action Arbitration}\\ \text{PTO - Public Service Center PTO Publications}\end{Bmatrix}$$

Figure 3.7 a) A design matrix showing the transition from the patent system functional to physical space of the design. A check mark indicates a relationship exists between the associated functional requirement and physical parameter. b) A design matrix showing the transition from the physical to the process space.

First, figure 3.7a represents the transformation of the functional space into a physical design in the form of laws that fulfill the patent system functional requirements as described earlier in figure 3.5. A check mark in the matrix represents a strong interaction between the associated FR's and PP's. A zero represents a weak relationship between FR's and PP's. In this first matrix, the physical design of the statutes and policies maintain the independence of functional requirements. For example, there is nothing stated in the laws governing patentability that would effect the ability of the system to either protect the exclusive rights of the patent or to encourage innovation in the general public.

Figure 3.7b represents the transformation of the physical parameters (i.e., physical laws) into an implemented design through a set of process variables. The process variables are described in the statutes and policies of the patent system. The patent examiner has the primary responsibility of determining if an application meets the requirements of invention described in the patent law. If the patent examiner can not make a determination satisfactory to the customer(i.e., inventor), then the Board of Appeals can be invoked to render a decision on the patent application.

In addition to the patent examiner and the Board of Appeals, a new program has been implemented to reinforce the quality of patents issued. The program is a coordinated effort between the patent examining corps and the American Intellectual Property Law Association (AIPLA). The goal of the program is to determine how current examining practices can be improved to increase the certainty that patents issued are valid as well as to improve the appearance and record of the patents.

The issue of the quality of patents is a major source of coupling between the requirements to grant patents and those to protect the rights of patent owners. For example, if a patent is granted for an invention that does not meet the novelty requirement for patentability, it may potentially be infringing on another patent. In fiscal year 1987 there were over 137 thousand applications filed and over 88 thousand patents were issued according to the PTO annual report. The patent system has accumulated millions of patents, all of which must be checked to determine the novelty of new applications. This problem of successfully checking the novelty of an application, which is due to coupling between patent system functional requirements, will only magnify each year as more and more patents are issued. The automation of the patent filing system now underway at the PTO is a positive step towards reducing such coupling.

On December 12, 1980 public law 96-517 was passed which amended sections 301 and 302 to the existing U.S. Code, title 35 in an effort to decouple the requirements of granting patents and protecting them. The new law allowed for the reexamination of issued patents based on new evidence indicating prior art existed at the time of invention. This new evidence may be reported anonymously by any individual at any time to the PTO. The primary goal of this law was to reduce the number of infringement cases being litigated through the courts.

This new law had two substantial effects on the implementation process of the patent system. First, it essentially extends the examination process throughout the life of the patent. Second, it provides an alternative to seeking civil action in the courts or arbitration for a patent owner or other user of the invention cited in the reexamination request.

The second effect has a positive influence on minimizing the information content of the patent system by reducing the need for litigation in civil courts. In fiscal year 1987, the PTO received 240 reexamination requests of which 48 are known to be in litigation. Without this statute, all 240 cases may have been litigated through the courts. Of the 240 requests, 195 were found to be in need of reexamination, 33 patents were confirmed for all claims, 121 patents modified or added claims and 33 patents had all claims revoked. The issues concerning minimizing information will be examined in more detail in the next section.

The extension of the reexamination period throughout the life of the patent, although minimizing information content due to civil action taken on patents,couples the requirements of granting patents and of protecting patent rights. In a sense the patent is never really issued since it is subject to reexamination throughout it's 17 year life (14 years for design patents). Because of this process variable allowing reexamination, the functional requirement of granting a patent to inventors for their discoveries is not fulfilled.

The remedy for this problem requires the addition of a new variable that will maintain the independence of the functional requirements in the process space. The allowance of the reexamination process can be maintained as it exists in sections 301 and 302 of title 35. However, in order to fulfill the first requirement of issuing a patent, a set window of time must be defined (e.g., 5 years), after which a patent may no longer be subject to revocation.

Similarly, to protect an individual who had practiced the invention prior to the time of invention by the patent owner, a reexamination may be requested in order to protect said individual against infringement action. If prior art is shown through reexamination to exist prior to invention by the patent owner, the patent owner retains all rights to the patent if the five year window of time has elapsed. In addition, in such a case, the right to practice said invention by the person establishing prior use will be protected against infringement actions.

By designing these two process variables of a window of time after which a patent may not be invalidated and protection against infringement for prior users, the design matrix shown in figure 3.7b becomes uncoupled. In this design, a patent would be granted subject to reexamination and then after the invention has been commercially available for over five years it can be declared incontestable with respect to the novelty and nonobvious requirements for patentability.

3.5.3 Nonobvious Requirement

Figure 3.8 illustrates a design matrix for a more detailed part of the patent system design that must define the subject matter which is entitled to a patent.

Patentability Requirements Design Matrix

$$\begin{array}{ccc} \text{Functional Requirements} & \xrightarrow{\hspace{6em}} & \text{Physical Parameters} \end{array}$$

$$\begin{Bmatrix} \text{Novelty} \\ \\ \text{Nonobvious} \\ \\ \text{Utility} \end{Bmatrix} = \begin{bmatrix} \checkmark & \checkmark & \text{O} \\ \\ \checkmark & \checkmark & \text{O} \\ \\ \text{O} & \text{O} & \checkmark \end{bmatrix} \begin{Bmatrix} \text{Title 35, U.S. Code} \\ \text{Section 102} \\ \text{Title 35, U.S. Code} \\ \text{Section 103} \\ \text{Title 35, U.S. Code} \\ \text{Section 101} \end{Bmatrix}$$

Figure 3.8 A design matrix illustrating the relationships between the requirements for patentability of the U.S. Patent System and the statutes as they are written in the code of U.S. law.

The definition of nonobviousness and novelty stated in sections 102 and 103 respectively do not maintain the independence of the two functional requirements. In this particular situation, the two requirements are in fact redundant as was previously argued in section 3.3.2. If an invention can be shown to have utility in function and which is also shown to be novel, then logically it is therefor not obvious. If it were obvious, others would recognize the utility of it and subsequently make use of it.

In determining questions of obviousness, the process of innovation must be considered. When designing solutions to problems, individuals must organize information and knowledge in such a way that an invention is recognized by the inventor as an obvious extension of the prior knowledge. In rewarding an inventor with a patent, both the efforts required of organizing the information as well as the ability to extract a useful invention from the information must be considered. Because only the latter of these two efforts are considered when implementing the obviousness requirement for patentability, the patent system does not always encourage innovation.

This coupling results partly as a consequence of the process variables and partly because of physical parameters. In the process of judging obviousness, an expert witness is often called to testify. The expert is familiar with the knowledge in the field and the associated invention. If the invention does not seem obvious to an expert after presented with the information organized by the inventor, then the expertise of the witness should be questioned. If on the contrary, the witness finds the invention to be an obvious extension of prior art, the patent should not be invalidated based on this testimony.

Because obviousness is a quantity of varying degrees that is difficult to judge or measure, one suggestion would be to remove it completely because of the implementation problems involved. The nonobvious requirement can however be judged objectively if based on a set of more concrete rules that have measurable or determinable quantities of reasonable certainty. Such rules are suggested below.

1. For obviousness to be determined as grounds for refusal of granting a patent in cases where the utility and novelty of the invention have been shown, evidence that one or more inventors discovered the same invention at the same time or within a reasonable time thereafter must be shown.
2. In such cases, the individual inventors making the same discovery may all retain equal rights to license or arbitrate a deal among themselves for the retention of patent rights.

3.6 Application of the Information Axiom

Information content in the context of the patent system design is defined as the amount of knowledge or expertise that must be applied to the system in order for it to perform as intended by congress. Minimizing information content in the patent system design means to reduce the uncertainty in the meaning of the law such that a minimal number of questions arise concerning the application of said law to a broad range of cases.

When difficult questions due to this uncertainty arise, the patent system requires an input of information in order to implement the design as intended. This information is supplied first by the Patent and Trademark Office and secondly by the judicial system.

3.6.1 Input from the Judicial System

Difficult patent cases require a lot of information to make a determination as to the validity of a particular invention. These most difficult cases tend to end up being decided by a judge in civil court. Judges are perhaps the least equipped to deal with the information needed to decide such cases. The information needed includes understanding the technology, the inventor and the innovative process, as well as the intent of congress in their design for the patent system.

On October 1, 1982 the U.S. Court of Appeals for the Federal Circuit was formed and given jurisdiction over all U.S. District Court decisions in patent infringement actions. Lee(1987) of the American Bar Association points out that approximately half of the Federal Circuit courts time is spent resolving patent infringement disputes due to the complexity of such cases. Furthermore, only three of the twelve appointed judges were "patent-experienced" before their appointments.

The judges of the Federal Circuit spend half their time in the role of patent system designers. It is their decisions that fine tune and adjust the accuracy of congressional design and PTO implementation of the patent system.

When a machine is not accurate enough for a particular application, the situation is remedied by employing an individual, knowledgeable in the workings of the machine, to operate it. By doing this, the knowledge of the operator compensates for the lack of machine accuracy. The less accurate the machine is, the more information that is required for it to perform with the desired accuracy, thus requiring a more skilled operator to implement or manufacture the design.

The patent system is a machine designed to provide a service to inventors through the encouragement of their innovative efforts, by securing for them the rights to their discoveries. When the patent system design (i.e., laws) is not accurate enough to determine the patentability of a particular application, a skilled operator is needed to supply the necessary information to the system in order to implement the system as designed by congress.

Judges are the "skilled operators" assigned the task of supplying the necessary information to the patent system so that the intent of the congressional design is realized. The question that surfaces from this perspective is: are judges prepared with the necessary skills to supply the required information to the system? Judges must understand the patent system in purpose and construct, they must understand the process of innovation and the needs of the customer (i.e., the inventor), and they must have experience in some field of technology. These should be the minimal requirements for any one presiding over patent infringement cases.

To find judges with the necessary skills may be difficult, but it is one possible solution to the patent system implementation problems. Implementing infringement proceedings may still be difficult within the judicial system and a solution outside the system would be another alternative.

Rather than employing the civil courts in cases of patent infringement, one proposition would be that a specialized group of individuals skilled in the art of technology and patent law be employed to make decisions on patent infringement cases.

The general meaning of information content in the context of the patent system has been described here as the decisions of the courts on infringement actions. The PTO(1988) reported that the U.S. Court of Appeals for the Federal Circuit disposed 90 of the 161 patent cases brought before them in fiscal year 1987 and left 71 cases pending.

These cases represent less than one percent of all patent applications acted on by the PTO, but the cost of these cases represents a considerably greater percentage of the total patent system costs. More efforts should be made towards increasing the certainty in meaning of the statutes, thus decreasing the information content requirements of the patent system and minimizing the need for litigation in the courts. This is perhaps the most difficult challenge for the patent system designers.

3.6.2 First-to-file system

The majority of the world community grants patents to the individual who is the first to submit an application for patent. The current U.S. system grants patents to the individual who is the first inventor. The current efforts of the PTO to harmonize the U.S. patent system with that of the rest of the world suggests that the U.S. system be altered to a first-to-file system.

The merit of the first-to-file system is that it reduces the information content required to be input into the system design in determining patent ownership disputes. Proving who the first inventor is requires much more information and effort than does showing who the first to file an application is. This would be a viable way to reduce information content except that it violates the requirement of granting patents to **inventors** for their discoveries.

According to Webster(1987), to invent is to bring something new into existence. The constitutional mandate for the patent system declares that the exclusive rights of inventors to their discoveries be secured for a limited time. To go to a first-to-file system would not be in the spirit of the constitution. The rights of the inventor are not secured, but rather protection is given to those capable of submitting applications the fastest. This would favor large corporations with extensive legal and patent departments and would be deleterious to the individual inventor or small company.

The constitution should not be repealed in this situation and the U.S. should retain the first-to-invent system as well as encourage the international community to adopt this system.

3.7 Summary of Patent System Findings and Recommendations

3.7.1 Based on the Independence Axiom

1. By incorporating new standards of anticompetitive activity into its rules, the doctrine of misuse is coupled to the antitrust statutes.
   The Congress should reform the "doctrine of misuse" with legislation that declares misuse of patent rights only when there is a direct violation of the antitrust statutes.

2. The promotion of free trade is coupled with the functional requirements for the protection of intellectual property. Counterfeiting and other violations of intellectual property rights is an international problem that will have increasing importance for the United States as the world market for American goods expands.
   The solution to this problem will come about only through international cooperation, treaties and laws. The U.S. should be a leader and push forward with determination in the effort to develop international consensus concerning the protection of intellectual property rights. Recent efforts by the PTO, WIPO and GATT should be aggressively pursued.

3. The patent system should not expend efforts in programs such as PROJECT XL since it is not part of their mandate to develop educational programs to encourage innovation in children. This program has merit and should be continued and supported by the Department of Education.

4. The quality reinforcement program initiated by the patent examining corps of the PTO and the American Intellectual Property Law Association is designed to study the coupling problem between the functional requirement to grant patents and the one to protect patent rights. This program is a positive step toward decreasing the uncertainty concerning the validity of patents and should be staunchly supported.

5. Sections 301 and 302 of title 35, U.S. Code, grants the right to any individual to request for a reexamination of an issued patent at any time. Since a patent is always subject to reexamination, the patent owner is never granted exclusive rights beyond reproach, and thus the functional requirement for granting patents is never fulfilled.
   The patent system should adopt a standard length of time, after which the patent owners rights to the use of the

invention are incontestable. However, any individual who can show prior use after the patent has been declared incontestable will retain the rights to use the patented subject matter.

6. The patent system must incorporate the needs of the customer (i.e., the inventor) into its design. The design process requires the organization of information in such a way that the solution to a problem becomes obvious to the inventor. The requirement of nonobviousness does not recognize this process in its current design.
   The requirements for nonobvious subject matter should be based on the following two fundamental objective rules.

   a. For obviousness to be determined as grounds for refusal of granting a patent in cases where the utility and novelty of the invention have been shown, evidence that one or more inventors discovered the same invention at the same time or within a reasonable time thereafter must be shown.
   b. In such cases, the individual inventors making the same discovery may all retain equal rights to license or arbitrate a deal among themselves for the retention of patent rights.

3.7.2 Based on the Information Axiom

1. Given the level of information content required for the patent system to function, highly skilled individuals in technology and law are needed to supply information to keep the system running smoothly. Currently the system employs judges to supply such information to the most difficult patent implementation problems. Because the majority of judges lack strong qualifications in both technology and patent law, their input is very inefficient.
   The most difficult patent cases should be handled by specially trained individuals with relevant experience in both technology and law. This could be either a specialized court or a specialized jury of individuals.

2. Adopting a first-to-file system is not a viable way to reduce information content because it violates the fundamental functional requirement of the patent system of granting patents to inventors for their discoveries.

The U.S. patent system should retain the first-to-invent system and encourage the international adoption of such a system as well.

Appendix 3A - The Laws of the U.S. Patent System Continued

3A.1 Part III - Patents and Protection of Patent Rights

This third part of Title 35 fo the U.S. Code governing patents defines the extent of and procedure for the protection of patent rights. Section 251 through 256 describe the details of actions to be taken in cases where amendments or corrections to patents are required. The purpose of these sections are to make allowance for any mistakes that the applicant or attorney therefore may have made on the application that would subsequently result in an invalid patent.

Sections 261 and 262 of this title assigns the attributes of personal property to a patent so as to allow the owner of said patent to license, sell or surrender freely all patent rights. Sections 271 and 272 define the act or acts that constitute infringement of patent rights and would subsequently entitle the patent owner to equitable retribution for said infringement.

To find that a patent has been infringed upon, it must be shown that the invention was made, used or sold during the time of the patent by someone without due authority. In October of 1984 Congress amended section 271 (P.L. 98-622) to add to the exclusive rights provided by a patent to make the import of any products produced outside the United States with a process protected with a U.S. Patent to be an infringement of said patent. Furthermore, with the passage of the Omnibus Trade and Competitiveness Act of 1987, in a case involving the import of a product allegedly produced by a patented process, the burden of proof rests with the alleged infringer to demonstrate that the product was not produced with a process protected by a U.S. Patent.

Sections 281 through 294 are grouped together in this title as the "Remedies for Infringement of Patent and Other Actions." Section 281 guarantees the right of the patentee to pursue civil action in cases of infringement. Section 282 declares that in any infringement case the patent claims shall be assumed valid and the burden of proof rests with the party asserting invalidity. Also required under this section is that the party claiming invalidity or non-infringement of patent claims provide notice to the patent holder at least thirty days before the trial.

Section 283 asserts the privilege of the court to grant injunctions to prevent violation of patent claims. Section 284 permits the courts to assess damages for patent infringement cases at no less than a reasonable royalty for the use made by the infringer and for as much as three times said amount. The allowance of treble damages is at the discretion of the court and is designed for the purpose of discouraging willful and knowledgeable infringement of patent claims.

Section 285 permits the court to award reasonable attorney fees to the prevailing party. The purpose of this section is to prevent undue prolongation of litigation by the loosing party's counsel that result in unjustified high attorney fees charged to the just party. It is at the discretion of the court to determine when such awards are warranted.

Section 286 provides a six year statute of limitations after which no recovery can be had for prior infringement. The purpose of this section is to encourage patentees to present claims for infringement promptly so as to allow the alleged infringer to correct its mistake or otherwise resolve the perceived problem.

Section 287 requires that patent holders mark the patented subject matter so to prevent an individual from infringing unknowingly on the patent claim. Section 288 provides a patentee to file a disclaimer to invalidate a specific claim of the invention that was subsequently determined to be invalid without penalty for infringement. The purpose of this section is to eliminate the common law practice that if a patent was in part not valid, then it would be completely invalid.

Section 289 defines an additional liability for those found guilty of infringing upon a design patent to the extent of the total profit gained from said infringement. Section 290 provides for the notification to the Commissioner of the PTO of all cases filed in the courts under this title within one month of such action.

Section 291 affords the owner of a patent the right to take civil action against another alleged to have an interfering patent. Section 292 declares the false marking of a non-patented object as patented and offence liable for a fine of not more than $500. The purpose of this section is to protect a patentee from fraudulent use of the patent name and to protect the public from deception by advertising the existence of a invalid patent.

Section 293 is designed to insure that nonresident patentees are served notice of any proceedings affecting their patent rights by designating the name and address of a person residing within the U.S. on whom notice can be served.

Section 294 makes allowance for arbitration as an alternative to civil action through the judicial system for the settlement of patent disputes. This section became effective in February of 1983 after the amendments incorporating the arbitration provision into the laws governing the PTO (P.L.97-247) were passed by the House and Senate in June and August of 1982 respectively. The advantages of the arbitration amendment were described in a House Report by the Judiciary Committee (No.97-542) on May 17, 1982.

> The advantages of arbitration are many: it is usually cheaper and faster than litigation; it can have simpler procedural and evidentiary rules; it normally minimizes hostility and is less disruptive of future business dealings among the parties; it is often more flexible with regard to scheduling of times and places of hearings and discovery devices; and, arbitrators are frequently better versed than judges and juries in the area of trade customs and the technologies involved in these disputes.

The purpose of the arbitration is to enhance the patent system with the above stated advantages and secondly to relieve some of the burden of the judicial system involved in litigation disputes. In 1987 there were 116 new court cases involving patent disputes, 131 patent cases disposed of and 71 cases pending at years end.

Section 301 of this title declares the right of any citizen to provide the PTO with evidence of existing prior art in the form of a prior patent of written publication. Furthermore, the confidentiality of identity for the individual providing the evidence is guaranteed. The purpose of this section is to encourage the introduction of evidence for patent infringement that would otherwise not be brought forth.

The final sections of the third part of this title defines the allowance of and procedure for the reexamination of patents by the PTO. The purposeful thrust was to provide for increased certainty in the validity of patents. By providing for the reexamination of doubtful patents by the PTO based on new evidence, expensive litigation proceedings are intended to be obviated.

3A.2 Part IV - Patent Cooperation Treaty

This final part of Title 35 was enacted by the Senate and House in June and November of 1975 respectively for the purpose of implementing the Patent Cooperation Treaty. This treaty resulted from a U.S. initiative in 1966 to determine a means for reducing the duplication of efforts associated with filing and processing patent applications in more than one country.

Sections 361 through 368 of this title define the international interface between the U.S. patent system and that of all other countries which conform to the Patent Cooperation Treaty. This includes the establishment of a Receiving Office for international applications, and International Searching Authority and a procedure for the processing of international applications. International applications filed in the U.S. and subsequently refused an international filing date by a foreign authority are subject to review by the Commissioner of the U.S. PTO upon request of the patent applicant. The provision for secrecy described in section 181 of this title will also be upheld for any international application that contains information for which its secrecy is vital to the security of the United States.

Section 371 through 376 govern the national operation of the Patent Cooperation Treaty. This includes the handling of patents, requirements and procedures for the PTO in concurring with the treaty, the publication of patents and the designation of fees payable to the U.S. PTO.

3.8 References

Computer and Business Equipment Manufacturers Association, "The Computer, Business Equipment, Software, and Telecommunications Industry 1960-1995", Global Engineering Documents Inc., 1986.

Court of Customs & Patent Appeals, "Application of Chatfield", United States Code Annotated, West Publishing Co., St. Paul, MINN., 1984, Note 221, p.35.

F.2d - Federal Reporter, Second Series, Cases argued and determined in the U.S. Court of Appeals, U.S. Court of Claims, U.S. Court of Customs and Patent Appeals and Temporary Emergency Court of Appeals, West Publishing Co., St. Paul MINN.

Hill, Eileen, "Commerce Department Program Seeks Greater Protection for U.S. Intellectual Property Rights", Business America, March 18, 1985, p.4.

Lee, William Marshall, Hearing before the Subcommittee on Patents, Copyrights and Trademarks, Committee on the Judiciary United States Senate, 100th Congress, S.Hrg.100-213, February 17, 1987, pp.126-148.

Leslie, Application for Patent, United States Code Annotated, West Publishing Co., St. Paul, MINN., 1984, Note 232, p.354. (See also 547 F.2d 116 for full description of case)

Patent and Trademark Office, Annual Report for Fiscal Year 1987, U.S. Department of Commerce, January 1988.

Quigg, Donald J., Speech to the Second Annual Conference on Global Competition, Rensselaer Polytechnic Institute, Albany, New York, June 25, 1987.

Susman, Thomas M., Gerlach, Wendy R., "The Need For Legislative Reform of the Law of Patent Misuse", Hearing before the Subcommittee on Patents, Copyrights and Trademarks, Committee on the Judiciary United States Senate, 100th Congress, S.Hrg.100-213, February 17, 1987, p.155.

U.S. - United States Reports, Cases Adjudged in the United States Supreme Court, U.S. Government Printing Office, Washington, DC.

Webster's Ninth New Collegiate Dictionary, Marriam-Webster Inc., 1987.

CHAPTER IV - SUMMARY AND CONCLUSIONS

4.1 Summary of Thesis

The principles underlying the design process are the same for the design of an electric circuit as for the design of the patent system. Although the methods used to aid in designing a circuit may differ from those used in developing a system of laws and policies, the principles or axioms that use the knowledge formalized through such methods are universal in all contexts. That such axioms exist is the primary proposition of this thesis. This proposition has been tested and analyzed in three ways using the design axioms uncovered by Suh(1988).

First, Chapter I introduced the two design axioms (i.e., Independence and Information Axioms) that are the focus of this inquiry and placed them in the broader framework of a design philosophy or science. The Science of Design envisioned here is composed of three main components; (1) the study of knowledge representations, (2) the study of knowledge formalizations, and (3) the study of the social context of the design process. The axioms presented here contribute to the first two of these components, while an axiom or axioms that describe the social context of the design process have not yet been postulated. Thus, the axioms of design constitute only a subset of the principles governing the design process.

The second test for these axioms is accomplished in chapter II by comparing them with the methods of design developed and proven in engineering contexts by Taguchi(1987). If Taguchi's methods are proven for engineering design, then their application should be consistent with the basic principles set forth in the axioms. It was shown that Taguchi's use of statistical sum-of-the-squares metric to measure the influence of the various control factors on the output response is applied to choosing adjustment factors based on the independence axiom. The sum-of-the-squares metric is used as a measure of functional dependence in which adjustment factors are chosen so to maintain the independence of functional requirements.

The signal-to-noise metric is shown to be a measure of information content. This metric is used to optimize the filter circuit in Chapter II by tightening the design range, thus minimizing the information content. Thus, the application of the sum-of-the-squares and signal-to-noise metrics are grounded in the fundamental principles of the independence and information axioms.

The third and final test for these axioms is conducted in Chapter III by applying these principles of design to the system of laws and policies that make up the United States Patent System. If the axioms are true, then they should be applicable to design in all contexts, whether designing a circuit, law, policy, organization or any other object. Chapter III develops a design representation for the patent system from which the application of the design axioms towards the recognition and answering of important policy questions is facilitated.

The findings and recommendations reported in Chapter III that were based on the application of the design principles are representative of some of the current policy questions concerning intellectual property rights in 1988. This suggests that the utility of the design representation developed in Chapter I and applied to the patent system in chapter III is that it helps lead the designer towards asking the right set of questions. The forming of the wrong problems and the posing of the wrong questions are often the source of design failures. The principles of design presented here are meant to aid in forming the right questions and determining the actual problems of design.

4.2 Engineering Design Compared to Policy Design

There are some important distinction to be made here between designing engineering systems and designing policy or legal systems. It should not be implied from this thesis that these two are the same in context or that design engineers should replace current law and policy makers, but rather that there is a common set of principles that guide both engineers and policy makers in their designs. Although engineering systems designers would certainly offer a unique perspective to policy and law design, this is not the main argument presented.

When designing engineering systems, if the physical parameters chosen to achieve the conceived functional requirements are not manufacturable with the current process, the process itself may be altered, within cost and other constraints. However, when designing policies, if one finds the process (i.e., Democratic Process) to be inadequate in achieving a chosen physical parameter (i.e., law or policy), altering the democratic system outlined in the Constitution is much more difficult, and not necessarily desirable. The alternatives in this situation are to either define a new set of functional requirements or to search for another physical parameter(i.e., law) which will achieve the same functional requirements and will be implemented with the given process.

When designing laws and policies there is seldom a clear relationship between the functional requirements, physical and process parameters which can be precisely or mathematically expressed. This makes methods used in engineering contexts, such as the semangularity and reangularity metrics in Chapter II, sometimes impossible to utilize for designing law and policies.

Also, because it may not be possible to experiment with policies the way one can experiment with a machine or non-human object, the utilization of Taguchi methods may not be possible in the context of designing policies and laws. However, by organizing a system of laws and policies into a design representation with functional requirements, physical and process parameters and qualitatively studying the relationship between these parameters, the elemental principles of design can be applied just as they were in Chapter II for a design in the engineering context.

4.3 Implications for Design Education

The design philosophy described in this thesis is one in which a continuing transition between knowledge in the functional, physical, and process spaces takes place as one proceeds in a hierarchical nature from the most broad and fundamental functional requirement to the most detailed physical and process parameters of the design. An important implication for the characteristics of a good designer is the ability to think and process information in these three different "design spaces". The knowledge and culture of the functional space is most often different from that in the process space. Therefore, a designer must be capable of understanding and merging these spaces if a manufacturable product or an implemented policy or law is to emerge.

For example, the disparity in knowledge and culture between marketing groups that often define the functional requirements of a particular design and the manufacturing groups that must choose the process variables in order to produce a product are often quite different. There are also disparities in knowledge and culture between the legislature that enacts laws and the courts that implement them. To assist in this transition between the functional, physical and process spaces, it has been proposed here that there are fundamental principles that govern the design process which will help the designer to determine the feasibility of a design, as well as to choose between alternatives.

The disparities of knowledge and culture represent barriers to the integration of the design process among the various contributors (e.g., marketing, engineering, manufacturing, etc...). The development of a design science with a fundamental set of general principles governing the design process could aid in overcoming such barriers. The principles of design can be taught in many contexts to students of various disciplines and should have a unifying influence among the disciplines.

It was in the interest of progress that broad areas of study were modularized into specialized disciplines. The motivation behind this was that two individuals could make more progress by dividing their subject area into two disciplines and pursuing each one independently, than by both pursuing the more broadly defined discipline together. The division of knowledge into smaller components is still today a useful way to further ones understanding of a particular discipline or field of study.

However, when one considers using these components in an integrated way, as is the case of most designs (that incorporate knowledge from several disciplines), an understanding of the relationships between these disciplines becomes as important as the disciplines themselves. The new discipline proposed here as design science is in part the study of the relationship between various disciplines of knowledge and how they interact to form a design.

By studying the interaction between disciplines during the formation of a design, the individual disciplines will be advanced by broadening their base of understanding. Just as when constructing a building, in order to progress higher, a broader base is required, so too when advancing a discipline, in order to progress, a broader base of knowledge on which to build is necessary. This is the dilemma of individuals whose education is too narrowly focused on a single sphere of knowledge or discipline. That is, in order to advance the discipline in knowledge and understanding, the individual must have a solid and wide base to build on.

The necessity of having this broad base of understanding is particularly crucial to the advancement of design science. All design (e.g., engineering, political, etc...) has some connection to a human need as its purpose. In some instances this may be obvious (e.g., civil rights laws, medicines, etc...) and at other times it may not be so obvious (e.g., nuclear weapons, circuit designs, etc...). Individuals involved in the design process take on the responsibility for satisfying human needs through their innovations in laws, policies, machines and all other engineered artifacts. It is not good enough, nor is it wise, to ignore this human purpose in the process of design. For this reason a designer must have a broad base of understanding and a strong facility for communicating across disciplinary boundaries. Further development of a design science and its elemental principles will serve to advance our ability to fulfill basic human needs through the design process.

> The specialist who is trained but uneducated, technically skilled but culturally incompetent, is a menace.
>
> David B. Truman,
> Address in Chicago to Columbia University
> Alumni, 15 April 1964.