WEAVING TIME INTO SYSTEM ARCHITECTURE:

NEW PERSPECTIVES ON FLEXIBILITY, SPACECRAFT

DESIGN LIFETIME, AND ON-ORBIT SERVICING

by

Joseph Homer Saleh

Ingénieur diplômé de l'École Nationale Supérieure de l'Aéronautique et de l'Espace (SUPAERO), Toulouse, 1997

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of
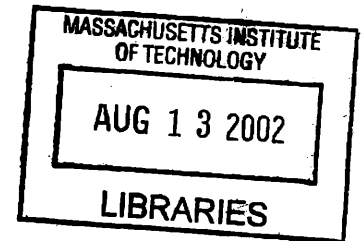
Doctor of Philosophy

in Aeronautics and Astronautics

at the

Massachusetts Institute of Technology

February 2002

[June 2002]

Author...
Department of Aeronautics and Astronautics

Certified by ....
Daniel Hastings, Professor of Aeronautics and Astronautics and Engineering Systems
Associate Director of Engineering System Division, Thesis supervisor
Director, MIT Technology and Policy Program

Certified by ....
Dava Newman, Associate Professor of Aeronautics and Astronautics
MacVicar Faculty Fellow. Thesis Committee Member

Certified by ....
Joyce Warmkessel, Senior Lecturer, Department of Aeronautics and Astronautics
Thesis Committee Member

Certified by ....
Triantaphyllos Akylas, Professor of Mechanical Engineering
Thesis Committee Member

Certified by ....
Wallace Vander Velde, Professor Emeritus of Aeronautics and Astronautics
Chairman, Department Graduate Committee

WEAVING TIME INTO SYSTEM ARCHITECTURE:
NEW PERSPECTIVES ON FLEXIBILITY, SPACECRAFT
DESIGN LIFETIME, AND ON-ORBIT SERVICING

by

Joseph Homer Saleh

Submitted to the Department of Aeronautics and Astronautics

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Aeronautics and Astronautics

Abstract

A roadmap for a comprehensive treatment of issues of flexibility in system design is developed that addresses the following questions: 1) What are the characteristic features of flexibility in system design? Can one clearly and unambiguously characterize flexibility, and disentangle it from closely related concepts? 2) What drives the need for flexibility in system design, and what are the attributes of an environment in which flexible designs should be sought and fielded? 3) How can one embed flexibility in a system design? 4) What are the trade-offs associated with designing for flexibility? What is the value of flexibility and what are the associated penalties (cost, performance, risk, etc.), if any? These are the fundamental questions around which this thesis revolves.

The first part of this work addresses the first two questions: Flexibility of a design is here defined as the property of a system that allows it to respond to changes in its initial objectives and requirements—both in terms of capabilities and attributes—occurring after the system has been fielded, i.e., is in operation, in a timely and cost-effective way. It is argued that flexibility should be sought when: 1) the uncertainty in a system's environment is such that there is a need to mitigate market risks, in the case of a commercial venture, and reduce a design's exposure to uncertainty in its environment, 2) the system's technology base evolves on a time scale considerably shorter than the system's design lifetime, thus requiring a solution for mitigating risks associated with technology obsolescence. In other words, flexibility reduces a design's exposure to uncertainty, and provides a solution for mitigating market risks as well as risks associated with technology obsolescence.

One way flexibility manifests its criticality to systems architects is in the specification of the system design lifetime requirement. The second part of this work addresses issues of design lifetime, and ways to provide and value flexibility in the particular case of space systems. First, it is shown that design lifetime is a key requirement in sizing various spacecraft subsystems. Second, spacecraft cost profiles as a function of the design lifetime are established and a cost per operational day metric is introduced. It is found that a cost penalty of 30% to 40% is incurred when designing a spacecraft for fifteen years instead of three years, all else being equal. Also, the cost per operational day decreases monotonically as a function of the spacecraft design lifetime.

An augmented perspective on system architecture is proposed (diachronic) that complements traditional views on system architecture (synchronic). It is suggested for example that the system's design lifetime is a fundamental component of system architecture although one cannot see it or touch it. Consequently, cost, utility, and value per unit time metrics are introduced and explored in order to identify optimal design lifetimes for complex systems in general, and space systems in particular. Results show that an optimal design lifetime for space systems exists, even in the case of constant expected revenues per day over the system's lifetime, and that it changes substantially with the expected Time to Obsolescence of the system and the volatility of the market the system is serving in the case of a commercial venture. The analysis proves that it is essential for a system architect to match the design lifetime with the dynamical characteristics of the environment the system is/will be operating in. It is also shown

that as the uncertainty in the dynamical characteristics of the environment the system is operating in increases, the value of having the option to upgrade, modify, or extend the lifetime of a system at a later point in time increases depending on how events unfold.

On-orbit servicing provides a way to physically access, upgrade, and modify a spacecraft. In other words, on-orbit servicing provides flexibility to space systems. A new perspective on on-orbit servicing is developed that focuses on the value of servicing as seen from the customer perspective. This contribution is based on three main ideas: The principal idea consists of estimating the value of servicing independently from its cost or specific implementation. The second idea lies in the observation that on-orbit servicing provides flexibility to space systems. The third idea recognizes that the value of servicing, contrary to what has been implicitly assumed by traditional approaches, is not limited to cost savings. Instead, it is shown that the value of flexibility provided by on-orbit servicing is an important component of the value of servicing. A valuation tool that leverages the advantages of Decision-Tree Analysis and Real Options is developed that captures this value of flexibility. Finally, while the results obtained are promising for the future of on-orbit servicing, this new perspective does not provide an argument for or against on-orbit servicing. Instead, it suggests a careful valuation process that focuses on the customer. Ultimately, a customer would opt for on-orbit servicing if the value of servicing a spacecraft exceeds the cost of doing so.

Thesis Supervisor: Daniel E. Hastings

Professor of Aeronautics and Astronautics and Engineering Systems
Director, MIT Technology and Policy Program, Thesis supervisor
Associate Director of Engineering System Division

To Homer, Maggie, Louis, and Philippe Saleh,

the thought of whom never leaves me.

To Karen Marais,

qui a été la moitié tacite de tout ce que j'ai écrit, et les deux moitié de beaucoup d'autres choses.

Acknowledgements

A sheer intellectual pleasure! I cannot find a better way to describe my experience at MIT. There are two tragedies in life, George Bernard Shaw is supposed to have said: The first is to fulfill none of one's dreams; the second is to fulfill all of one's dreams. In this sense, my years at MIT have been a "tragedy" in that I have achieved far more than I had dreamed of. For the many people who have contributed to my personal and professional/academic growth, it is with pleasure that I write the following acknowledgements. These few lines however are only the tip of the iceberg of my debt of gratitude towards them.

My first and foremost debt of gratitude goes to my co-advisor, **Professor Dava Newman**. The first semester at MIT was rather rough on me because of the lack of support structure and my financial situation. A week before I had planned to leave MIT for another university where I had a research assistantship, Dava offered me financial (and moral) support, convinced to stay here, and coached me on how to thrive at MIT. Ever since, she has been a most wonderful advisor, mentor, and friend. If anything good came out of my teaching and research at MIT, Dava has made it possible and I am immensely grateful to her.

I started working with **Professor Daniel Hastings** in September 1999 after he returned from his appointment at the Air Force. Dan provided the intellectual thrust from which this research sprung, and I am convinced this work could not have developed under any other guidance than his. While other faculty in the department dealing with space systems are occupied with short term deliverables, Dan can afford to ponder the future of aerospace, and tackle fundamental problems in the field. On a side note, Dan is the fastest reader on Earth (although I haven't timed him, I think he reads around 2-5 pages per minute) and I have greatly enjoyed our early conversations on flexibility. I am most grateful for the freedom he has given me in defining and pursuing my research without pressure or interference.

I have been very fortunate to have known and interacted with **Professor Earll Murman** and **Professor Eric Feron**. They have both been very kind to me and supportive, and I was very happy that they could make it to my defense. Eric always encouraged me to pursue further the quantitative aspects of my work (Chapter 5 on the optimality of the design lifetime is to a large extent the result of his encouragements). His comments and suggestions have significantly enhanced the quality of my work and I am very grateful for all what he has done for me.

Colonel Peter Young is an inestimable resource for the faculty, staff, and students of the Aero/Astro department. He is a wonderful storyteller who always enriches the often-simplified academic discussions on space systems with his wealth of knowledge and experience with real systems and management issues. I am very grateful to Pete for the many discussions we had (Chapter 4 on the impact of spacecraft design lifetime is a direct result of his advice), for his support and his willingness to help.

An old French idiom says all books are made of other books; a thesis is not any different. In the course of this work, I spent a lot of time with numerous **books and papers**. To these wonderful companions, I owe more than gratitude: almost a friendship.

Talking about friends, two of the most brilliant people I met at MIT are **Emilio Frazzoli** and **Cyrus Jilla**. While they have both made fundamental contributions to their respective fields, their technical skills are dramatically surpassed by their kindness and their support of other fellow students. I have been very fortunate to have met Emilio and Cyrus during my first semester at MIT, when I was having a tough time adapting to this new environment after several years spent in France, and looking for an

advisor and seeking financial support. They are wonderful friends who contributed tremendously to making my stay at MIT most rich and enjoyable. Emilio, Cyrus: Thank you very much for every thing.

Natasha Neogi is our resident genius and a very dear friend. She attended Oxford at the age of 14 and completed a couple of Masters degrees before joining the PhD program in the Aero/Astro department at MIT. Her eloquence, strong will, and technical skills are simply admirable and made every discussion we had most pleasant and enriching.

I have been blessed with many exceptional Lebanese friends. They have contributed to my personal growth far beyond my ability to communicate or ever pay back: **Walid Fayad** (the most "colorful" person according to his advisor, a genuinely warm, hospitable, and incredibly smart person), **Issam Bazzi** (the master of hidden Markov models, in speech recognition and life in general!), **Danielle Tarraf** (the iron lady with a heart of gold), and **Jihad Touma** (the most beautiful mind I have encountered and a wonderful astrophysicist/story teller). I am very honored to count **Mirna Daouk, Ziad Ferzly, Karim Karam, Nidal Khalil, Eugenie de Meuse, Maya Saîd, Ghassan Matar,** and **Ziad Sultan** and among my very dear friends.

Although English is my third language, my writing has evolved from a painful exercise to a most enjoyable one thanks to **Marsha Warren**. Her smile and (Mississippi) accent have always brightened my days at MIT.

I owe a special thanks to Professor Hastings' research group, **Elisabeth Lamassoure, Myles Walton, Annalisa Weigel,** and to my two consecutive officemates, **Sylvie Loday** and **Chris Carr**. They have been simultaneously my friends, colleagues, and sometimes, unknowingly, my mentors.

Finally, my greatest debt goes to my family, **Homer, Maggie, Louis,** and **Philippe Saleh**, as well as to **Karen Marais** (who is the smartest person I know). Their unconditional love and support is beyond my ability to describe and can never be compensated for. Thank you mom, dad, Loulou, Philo, Karen. This thesis is yours.

Contents

PART I: ON FLEXIBILITY IN SYSTEM DESIGN

7

PART II: SPACECRAFT DESIGN LIFETIME AND ON-ORBIT SERVICING

8

List of Figures

13

List of Tables

To The Reader

"I know that, despite my care, nothing will be easier than to criticize this [thesis] if anyone ever thinks of criticizing it. I think those who want to regard it closely will find, in the entire work, a mother thought that so to speak links all its parts. But the diversity of the objects I had to treat is very great, and whoever undertakes to oppose an isolated fact to the sum of facts I cite or a detached idea to the sum of ideas will succeed without difficulty. I should therefore wish that one do me a favor of reading me in the same spirit that presided over my work, and that one judges this [thesis] by the general impression it leaves, just as I myself decided, not by such and such a reason, but by the mass of reasons..."

Adapted from A. De Tocqueville. *Democracy in America.* 1835

Chapter 1

Introduction

1.1 Background and Motivation

The ephemeral nature of human life has been a major theme for philosophers, theologians, poets, and others, ever since the dawn of history. A myriad of human behaviors and artifacts (intellectual, artistic, even political institutions) stem from, or find the original impetus for their existence in an individual or collectivity's relationship with *time*. Just like "we are [physiologically] the children of gravity, we cannot see it or touch it, but it has guided the evolutionary destiny of every species, and has dictated the size and shape or our organs and limbs[1]", so are many of our psychological dispositions, behaviors, and constructs the children of our relationship with *time* (we cannot see it or touch it…) and the recognition of the transiency of human life.

Less profound but no less thought-provoking, is the transiency of human handiwork. Of all the structures and artifacts of antiquity, only an infinitesimal remnant survives today [Terborgh, 49]. Examples abound as well in more recent periods of industries, equipments, and products that exhibit an ephemeral relationship with *time*, and stand as modern reminders of the transiency of such artifacts. Typically a product progresses through a life cycle characterized by periods of growth, maturity, and decline, then it dies-out because of physical, functional, or economical degradation or inadequacy. At Cape Canaveral for instance, lie the remnants of the race to the Moon: concrete launch pads, bunkers, and steel gantries in ruins from the Mercury, Gemini, and

---

[1] D. Newman. "Human Spaceflight from MIR to Mars". AIAA-SF June 2000 Dinner Meeting. Sunnyvale.

Apollo missions. Similarly, outside Tucson, in the Arizona desert, one finds the Aerospace Maintenance and Regeneration Center[2] (AMARC), better known as the aircraft graveyard where over four thousand aircraft lie moldering in the sun (see Fig. 1-1). These modern ruins, familiar technological objects, stand as reminders that nothing is permanent. Through physical or functional degradation, or loss of economic usefulness, the hand of time lies heavy on the work of humans.
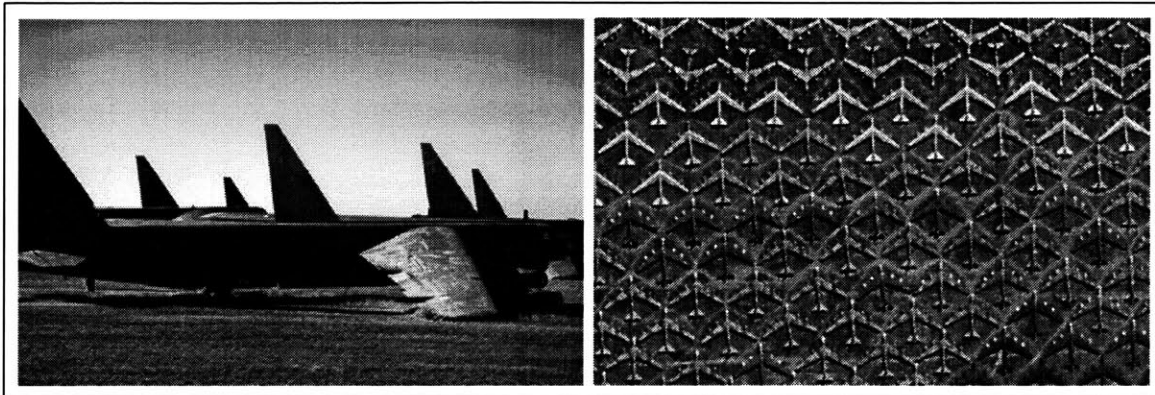
Fig. 1-1. B-52s moldering in the sun at the Aerospace Maintenance and Regeneration Center (AMARC), better known as the aircraft graveyard. To the right, a satellite image of a sector of the facility.

Several terms are used to describe this particular aspect of a product's relationship with time, namely the span of time from fielding a product to its retirement or replacement. These include "life span", "design lifetime" or the more recent term coined by Fine (1998), "clockspeed", to name a few. But what drives a system's life span? Terborgh (1949) in his seminal work *Dynamic Equipment Policy*, provides the following interesting discussion on unpredictability and changes in a product's environment, competition, and equipment's inexorable progression towards obsolescence and replacement:

> **Capital goods live out their mortal span in an atmosphere of combat, a struggle for life** [...]. Machines must defend themselves in a world where species spring up overnight, where **the landscape is never twice the same**, where the fitful winds of change are never stilled. [In the world of "capital goods"] death comes usually by degrees, through a process that may be described as **functional degradation**. It is a kind of progressive larceny, by which the ever-changing but ever-present competitors of an existing machine rob it of its function, forcing it bit by bit into lower grade and less valuable types of service until there remains at last nothing it can do to justify further existence.

---

[2] The AMARC provides storage, regeneration, reclamation, and disposal of aircraft and aircraft parts.

In other words, "machines", or **systems** in general, are **constantly faced** with both **changes** and unpredictability **in their environments**, as well as **functional aggression from competing products**. The **systems that thrive longer**, or have longer life span, are the ones that **are capable of coping with** unpredictability and **changes in their environment** (analyst's perspective). Conversely, **if a system is to be designed for an extended design lifetime**, the **ability to cope with** unpredictability and **changes has to be embedded in the system** (designer's perspective). These ideas are captured in Figure 1-2.
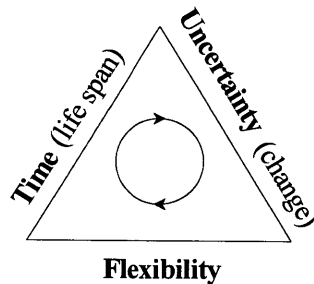
Flexibility

Fig. 1-2. The Trilogy: Time, Uncertainty, and Flexibility. Systems that have a longer life span are the ones that are capable of coping with uncertainty and changes in their environment (analyst's perspective). Conversely, if a system is to be designed for an extended design lifetime, the ability to cope with uncertainty and changes has to be embedded in the system (designer's perspective).

The discussion above is not without reminding us of another type of struggle for life and survival of the fittest[3]: that of biological species. In his *Origin of Species*, Darwin introduces the universality of the struggle for existence in the following terms:

> We shall now discuss in a little more detail the **struggle for existence**. The elder De Candolle and Lyell have largely and philosophically shown that **all organic beings are exposed to sever competition** [...]. Nothing is easier to admit the truth of the universal struggle for life [...]. I should premise that I use the term Struggle for Existence in a large and metaphorical sense, including not only the life of the individual, but success in leaving a progeny [Darwin, *Origin of Species*].

---

[3] This expression is not Darwin's invention, as often supposed, but that of his contemporary the philosopher H. Spencer: "I have called this principle by which each slight variation, if useful, is preserved by the term Natural Selection [...]. But the expression used by Mr. Herbert Spencer of Survival of the Fittest is more accurate and sometimes equally convenient." [Darwin, *Origin of Species*]

Given this omnipresence of the *struggle for existence*, the process of **Natural Selection operates "daily and hourly**, scrutinizing the slightest variations; rejecting those that are bad, preserving [...] all that are good": **Individuals or species that are better equipped to adapt to changing environments tend to be preserved longer**, according to Darwin's evolutionary theory.

The analogy with Terborgh's work is noticeable: Both living organisms and human artifacts, complex engineering systems for instance, strive in an ever-changing, competitively aggressive environment. Individuals or systems that are better equipped to adapt to changing environments, live longer, or outlive more *rigid* organisms or systems. The relationship between the inherent ability of an organism, or an inert system, to cope with changes and its life span is graphically illustrated in Figure 1-3.
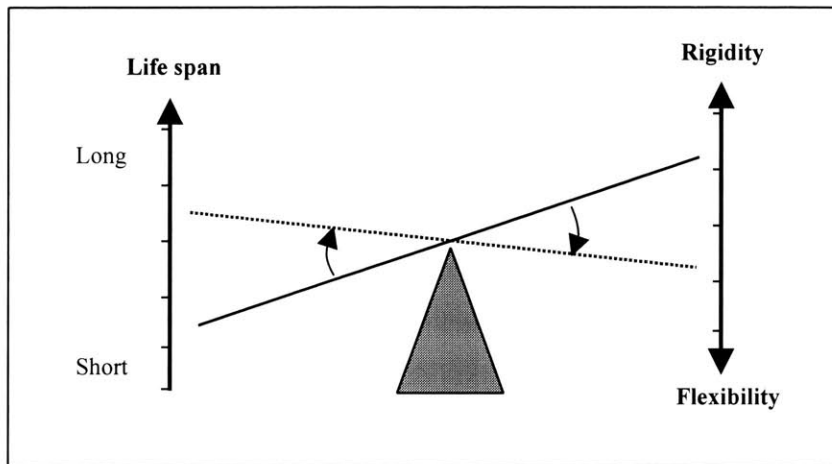
Fig. 1-3. Simple model relating a system's (living or inert) life span and its flexibility, which will loosely be defined for the time being as the ability of a system to handle changes.

Two separate observations make the investigation of the above-mentioned (and illustrated) relationship in the case of engineering systems a worthy endeavor:

1. On the one hand, current complex engineering systems are being designed for increasingly longer design lifetime. In recent years, several space programs for instance have chosen to increase their space segment design lifetime. Over the last two decades, communication satellites in the geo-stationary ring for instance have seen their design lifetime on average increase from seven to fifteen years. Similarly, a helicopter delivered today can exceed thirty years or 20,000 hours of operation. In most cases, increasing a system's design lifetime is driven by–and justified by–traditional economic analysis using Discounted Cash Flow

techniques such as the classical Net Present Value (NPV), and the desire to maximize the Return On Investment (ROI). However, extending a system's design lifetime has several side effects: Fielded systems with long design lifetimes can become obsolete, technically and commercially, before the end of their mission. In many cases, the initial circumstances from which the original system requirements were derived change or are modified during the system's operational lifetime. In the case of high-value assets, it is desirable to have systems that are flexible and can adapt to new or emergent missions and roles, instead of fielding new ones. Flexibility is thus a key property that should be embedded in high-value assets, particularly as they are being designed for increasingly longer design lifetime. But how can one design for flexibility? What are the design practices for embedding flexibility in design? What are the trade-offs associated with designing for flexibility (value of flexibility, cost penalty, performance penalty, etc.)?

2. On the other hand, flexibility has become in recent years a popular concept in many fields, particularly in most design endeavors. Indeed, for a multitude of disciplines, such as urban planning [McKinon, 88], architecture [Fox and Yeh, 99], finance [Trigeorgis, 96; Amram and Kulatilaka. 99], manufacturing [Raouf and Ben-Daya, 95], software design [Parnass, 79; Highsmith, 99] and others, flexibility is hailed as critical. However, few attempts have been made to formally and unambiguously define it. Intuitively, flexibility is understood as the ability to respond to change. Although essential, this feature nevertheless fails to distinguish it from other properties such as robustness. Furthermore, the literature on design is replete with terms related to a system's ability to handle change, such as adaptability, changeability, agility, elasticity, etc. But when one seeks to grasp their concrete content, such terms often fail. So what are the characteristic features of flexibility? How can one formally define it and quantify it?

The two above observations render the analysis of issues of flexibility, both from a conceptual and practical perspective, an exciting and challenging task. A comprehensive treatment of flexibility in system design should address the following questions:

1. What is flexibility? Can it be formally defined?
2. Why or when is flexibility needed in system design?
3. How can one design for flexibility? What are the design principles for embedding flexibility in system design?

4. What are the trade-offs associated with designing for flexibility? What is the value of flexibility, how can its value be quantified, and what are the penalties (cost, performance, risk, etc.), if any, associated with it?

These are the fundamental questions around which this thesis revolves. This section provided a conceptual background for the relevance of these questions, and drew an analogy between the need for human artifacts—capital goods or complex engineering systems—and biological systems to be able to adapt to an ever-changing and competitively aggressive environment in order to live longer or outlive more *rigid* organisms or systems. The following section is more specific as to the issues explored in this thesis.

1.2 Problem Statement

This thesis revolves around issues of flexibility in system design in general, and spacecraft design lifetime as well as on-orbit servicing as a means for providing flexibility to space systems in particular. A comprehensive treatment of issues of flexibility should address the four general questions mentioned in the previous section. It is however pretentious to attempt to address all these issues in a single thesis. Instead, this work addresses the first two questions mentioned above in the general case of system design, and explores the subsequent questions in the particular case of spacecraft design. More precisely, this thesis is divided into two parts: Part I addresses the two following questions:

1. What are the characteristic features of flexibility in system design? Can one clearly and unambiguously characterize flexibility, and disentangle it from closely related concepts?
2. What drives the need for flexibility in system design, and what are the attributes of an environment in which flexible designs should be sought and fielded?

One way through which flexibility manifests its criticality to systems architects is in the specification of the system design lifetime requirement. Part II addresses issues of design lifetime, and ways to provide and value flexibility in the particular case of space systems. In particular, the following questions are explored:

3. How do different spacecraft subsystems scale with the design lifetime requirement, and what is the total system mass and cost profile as a function of this requirement? It should

be noted that answering these questions is a prerequisite for addressing the following question.

4. Is there an optimal design lifetime for a system architecture, particularly spacecraft architecture? What does (or should) the customer ask the contractor to provide for a design lifetime, and why?

5. On-orbit servicing provides a way to upgrade or modify a spacecraft, or to extend its design lifetime. In other words, on-orbit servicing provides flexibility to space systems. What is the value of on-orbit servicing as seen from a customer's perspective? What is the value of flexibility provided by on-orbit servicing, and what are the appropriate valuation tools for capturing the value of flexibility?

The specifics of each Part and Chapter are discussed in the Thesis Outline below.

1.3 Thesis Outline and Contributions

This thesis is divided into two parts. Part I, On Flexibility in System Design, comprises Chapter 2, Extracting the Essence of Flexibility in System Design, and Chapter 3, The Case of Flexibility in System Design.

The purpose of Chapter 2 is to review the concept of flexibility as discussed in various fields of investigations, and to extract its characteristic features. In order to discuss any subject matter clearly, it is necessary to begin with a clear set of definitions. Indeed much can be gained through careful and consistent definitions of terms alone. Flexibility however is a word rich with ambiguity. Chapter 2 synthesizes a clear and consistent definition of flexibility, and to disentangle it from closely related concepts.

Chapter 3 identifies the situations and characterizes the environments in which flexibility in system design should be sought.

Part II, Spacecraft Design Lifetime and On-Orbit Servicing, comprises Chapters 4, On Spacecraft Design Lifetime, Chapter 5, Satellite Cost per Operational Day and Optimal Design Lifetime, and Chapter 6, Flexibility and the Value of On-Orbit Servicing: A New Customer-Centric Perspective.

Chapter 4 explores the impacts of the design lifetime requirement on spacecraft mass and cost to Initial Operating Capability (IOC). It first examines how different subsystems scale with the

design lifetime, then transforms these results to generate spacecraft mass and cost profiles as a function of the design lifetime.

Chapter 5 proposes to view in a system architecture the flow of service (or utility) that the system will provide over its design lifetime. It suggests that the design lifetime is a fundamental component of system architecture although one cannot see it or touch it. Consequently, cost, utility, and value per unit time metrics are introduced herein. A framework is then developed that identifies optimal design lifetimes for complex systems in general, and space systems in particular, based on this augmented perspective of system architecture and on these metrics. The analysis performed in this Chapter demonstrates two fundamental points: First that the optimal design lifetime changes as a function of the uncertainty in the dynamical characteristics of the environment the system is operating in. Second, that the value of having the option to upgrade, modify, or extend the lifetime of a system at a later point in time depending on how events unfold increases.

On-orbit servicing provides a way to physically access, upgrade, and modify a spacecraft. In other words, on-orbit servicing provides flexibility to space systems. A new perspective on on-orbit servicing is developed in Chapter 6 that focuses on the value of servicing as seen from the customer perspective, independently of any servicing architecture. This view is developed along with an appropriate valuation tool that captures the value of flexibility.

Chapter 7 contains the conclusions and recommendations for future work.

Although the chapters build on each other, they are nevertheless designed to be somewhat stand-alone. While this makes it easy for a person to randomly select and read one chapter from the thesis (and comprehend it), it nevertheless implies that there is a little overlap between each chapter. Hopefully this will not inconvenience the reader who wishes to read the whole thing in one sitting.

Asides from the particular contributions mentioned above, there are two conceptual contributions that are not detailed at any one point in the thesis; instead they are pervasive throughout the entire manuscript. In order to state the first one, let me first introduce an important terminology from linguistics. One of the pair of terms introduced by the Swiss linguist de Saussure in his seminal

work *Course in General Linguistics*[4] [Saussure, 15] are **synchrony** and **diachrony**, which together describe the two fundamental perspectives for the study of language. A synchronic perspective in linguistics refers to the state of a language as it exists at a given time. It is the study of a "snapshot" of a language that includes the analysis of structures and relationships within a language at a given time. Diachronic linguistics on the other hand analyzes the changes that have taken place over time in a language. It is a "cinematographic" study of a language that includes the identification and analysis of patterns and relationships in sounds, syntax, and vocabulary in the time domain [Finch, 00].

A fundamental conceptual contribution of this thesis is in the introduction of temporal considerations into system architecture. Using the terminology introduced above in the case of linguistics, this is equivalent to introducing a **diachronic perspective on system architecture,** when traditionally the synchronic approach (the "snapshot" approach) prevails. Indeed, system architecting has been traditionally viewed as a matching between two (vector) quantities, resources and system's performance. One approach fixes the amount of available resources and strives to maximize the system's performance; the other approach constrains the system performance and attempts to minimize the resources necessary to achieve the target performance [de Weck, 01]. The first approach operates with–and attempts to maximize–a performance per unit cost metric; the second approach seeks to minimize a cost per function (or performance) metric. This thesis views in a **system architecture** the **flow of service** (or utility) that the system will provide over its design lifetime. Consequently, cost, utility, and value per unit time metrics are introduced. It therefore suggests that we augment our understanding of system architecture by considering the system's **design lifetime**, as well as other time characteristics associated with a design, as **fundamental components of system architecture** although one cannot see them or touch them.

A second conceptual contribution of this thesis is in recognizing the fundamental relationships between **Time, Uncertainty, and Flexibility; the three faces of a same coin**. Time and uncertainty are intrinsically related, for if there were no tomorrow, there would be no uncertainty. Time transforms uncertainty, which in turn is shaped by the time horizon [Bernstein, 96]. Flexibility on the other hand reduces the exposure to uncertainty, thus allows a system the weather the heavy hand of Time.

---

[4] Ferdinand de Saussure, sometimes called the "father of modern linguistics", actually never published any work on the subject! After his death, his students collected his lecture notes and published them with the title *Cours de linguistique générale*, in 1915.

"In almost all textbooks, even the best, this principle is presented so that it is impossible to understand.
I have chosen not to break with tradition."

K. Jacobi on variational mechanics, *Lecture on dynamics*, 1843.

Part I

On Flexibility in System Design

Chapter 2

Extracting the Essence of Flexibility in System Design

"Complex problems have simple, easy to understand, wrong answers."

Anonymous.

"The splendid Philharmonic Orchestra in Berlin possesses a special quality
for which I can find no more appropriate expression than **flexibility**.
They have the capacity to adapt themselves to the dimensions of a Berlioz or a Liszt,
and of reproducing with equal mastery the variegated arabesques of the former and the thunderous
cannonades of the latter - yet they are able to exercise the restraint called for by the gentleness of a Hayden..."

P. Tchaikovsky.

The purpose of this chapter is to review the concept of flexibility as discussed in various fields of investigations, and to extract its characteristic features. In order to discuss any subject matter clearly, it is necessary to begin with a clear set of definitions. Indeed much can be gained through careful and consistent definitions of terms alone. Flexibility however is a word rich with ambiguity. While it is being increasingly used in various fields, few attempts have been made to formally define, quantify, and propose ways for achieving flexibility. This chapter proposes to fill in part this gap by synthesizing a clear and consistent definition of flexibility. It will do so by reviewing the usage of the term in various fields of inquiries, and show that it is indeed possible to clearly and unambiguously characterize flexibility, and to disentangle it from closely related concepts.

2.1 Flexibility: A Word Rich with Ambiguity

Flexibility has become in recent years a key concept in many fields, particularly in most design endeavors. Indeed, for a multitude of disciplines, such as urban planning [McKinnon, 78], architecture [Fox and Yeh, 99], finance [Amram and Kulatilaka, 99], manufacturing [Raouf and Ben-Daya (eds), 95], software design [Highsmith, 99] and others, flexibility is hailed as critical. However, few attempts have been made to formally and unambiguously define it. Intuitively,

flexibility is understood as the ability to respond to change. Although essential, this feature nevertheless fails to distinguish it from other properties such as robustness. Furthermore, the literature on design is replete with terms related to a system's ability to handle change, such as adaptability, changeability, agility, elasticity, etc. But when one seeks to grasp their concrete content, such terms often fail. One source of ambiguity therefore arises from the failure of the familiar characterization of flexibility, i.e., the ability to handle change, to distinguish it from other properties, particularly in the light of the proliferation of its pseudo-synonyms.

The following extract [Chen and Lewis, 99] is a good representative of this ambiguity where flexibility and robustness are used almost interchangeably, and in which "robust design [provides] flexible solutions":

> The robust design concept is extended to make decisions that are flexible to be allowed to vary within a range (called type II robust design)...The concept behind type II robust design for providing flexible solutions is represented below. For purposes of the illustration, assume that the performance is a function of only one variable x. Generally, in this type of robust design, to reduce the variation of response caused by variations of the design variables, instead of seeking the optimum value, a designer is interested in identifying the flat part of a curve near the performance target. If the objective is to move the performance towards M and if a robust design is not sought, then obviously $x = \mu_{opt}$ is a better choice. However for a robust design $x = \mu_{robust}$ is a better choice.
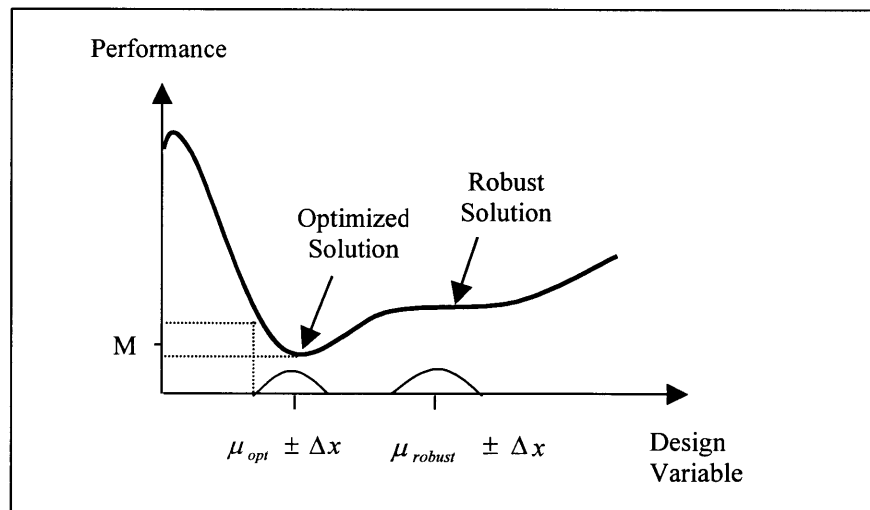
Fig. 2-1. Type II robust design: developing flexible solutions. Adapted from [Chen and Lewis, 99]. Variation of the design parameter around μ_{opt} causes greater variation in the performance than when the design parameter is set to μ_{robust}.

The above example sets the objective of "achieving flexible solutions" and equates that with (type II) robust solutions. Flexibility is thus turned into a by-product of the robust design methodology. It is arguable however whether such a discussion captures any distinctive feature of flexibility.

A comprehensive treatment of flexibility in system design should address the following questions:

1. What is flexibility?

2. Why or when is flexibility needed in system design?

3. How can we design for flexibility? What are the design principles for embedding flexibility in system design?

4. What are the trade-offs associated with designing for flexibility? What is the value of flexibility and what are the penalties (cost, performance, risk, etc.), if any, associated with it?

The literature of the different fields of inquiries mentioned above seldom addresses these questions holistically. Instead the focus is on one particular question at the detriment of the others, often the first and second question. The literature on Real Options is good example of this trend where the focus is primarily on capturing the *value* of flexibility. The following extract [Trigeorgis, 96] is the opening paragraph of a reference text on the subject; it illustrates the emphasis of the subject on the value of flexibility:

> Flexibility has value. While this statement is obvious at the conceptual level, it is surprisingly subtle at the applied level. Professional managers have long intuited that [flexibility is an important element] in valuation and planning decisions. But precisely how valuable is flexibility and how can its value be quantified?[1]

This chapter will focus on the first question: It proposes to review the concept of flexibility as discussed in different fields of investigations, and to extract its characteristic features. The

---

[1] Forward by Prof. Scott Mason, Harvard University, to *Real Options: Managerial Flexibility and Strategy in Resource Allocation*, by Lenos Trigeorgis, MIT Press, 1996.

objective is to synthesize a clear and consistent definition of flexibility, and to disentangle it from its pseudo-synonyms and related properties. The following questions will be addressed separately in the subsequent chapters.

The chapter is organized as follows: Section 2 provides a selected literature review of different fields of investigations that have addressed issues of flexibility. Section 3 proposes a definition of flexibility of a design, and carefully disentangles it from discussions on flexibility in the design process. A brief literature review of Robust Control and Robust Design is also provided and a definition of robustness is synthesized. Flexibility (of a design) and robustness are then contrasted, and a distinction is drawn as well between flexibility and universality of a design. Section 4 discusses three examples of flexible systems and the need for flexibility in system design, and illustrates the relationship between flexibility and a system's design lifetime. Section 5 touches on issues of flexibility in the context of distributed satellite systems. Section 6 contains the summary and conclusions.

2.2 Discussions of Flexibility: A Selected Literature Review

This section briefly reviews the various definitions of flexibility provided by three distinct fields of investigations: Flexibility in manufacturing systems, flexibility in multidisciplinary design processes, and real options thinking and managerial flexibility.

2.2.1 Flexibility in Manufacturing Systems

In the manufacturing community, different types of flexibility are defined based on the nature of *change* the production system can accommodate. The sheer amount of literature on Flexible Manufacturing Systems (FMS) is daunting. A great number of topics are addressed ranging from the design of manufacturing cells and machine grouping, to the scheduling, loading, and control of FMS [Raouf and Ben-Daya (eds), 95]. This section briefly reviews a handful of definitions among the numerous types of flexibility that are defined in this literature. **Volume flexibility** is defined as the ability of a production system to handle changes in daily or weekly volume of the same product, thus allowing the factory to operate profitably at varying overall production levels. **Product mix flexibility** is defined as the ability to manufacture a variety of products without major modification of existing facilities. **Routing flexibility** is defined as the ability to process a given set of parts on alternative machines. **Operation flexibility** is defined as the ability to

interchange the ordering of operations on a given part, thus allowing the ease of scheduling of its production [Suarez et al. 1991] [Taylor, 1991].

Flexibility in this environment is not only viewed as a reactive capability, it is also regarded as a competitive weapon which not only allows a company to *respond* to change, but also to create change and set the market pace for rapid production and innovation [Piore, 89].

Agility is another term related to the ability to respond to change. It was first introduced in manufacturing environments then broadened to encompass the extended enterprise. It is often loosely defined, and used to characterize different things in a business environment. For instance, in *Pathways to Agility*, Oleson (1998) describes "agile strategic planning processes", "agile automation", and discusses the need for "agile business relationships" with suppliers and customers. He defines agility as the "ability to respond with ease to unexpected but anticipated events[2]". Similarly, Fricke et al. (2000) define agility as the "property of a system to implement changes rapidly", and flexibility as the "property of a system to be changed easily and without undesired effects." "Agility" is thus used as a desired qualitative attribute for an enterprise to thrive in a hyper-competitive environment. It is difficult however to see how the definitions of flexibility and agility provided by Fricke et al. (2000) differ or overlap, and to grasp the concrete content of "agility".

2.2.2 Multidisciplinary Design and Flexibility in the Design Process

Current research has addressed the issue of *flexibility* in multidisciplinary design[3]. The focus of those efforts has been on achieving "flexibility in the design process." Typical approaches have consisted of incorporating designers' preferences with degrees of satisfaction in specifying design requirements. Thurston (1991) for example uses utility theory based *preference functions* to express designers' preference over single or multiple attributes. Wallace et al. (1996) define *specification functions* to indicate the subjective probability that performance levels are achieved. Mohandas and Sandgren (1989) recommend the use of *fuzzy goals* to model the degree of satisfaction level.

These approaches, along with others such as the interval methods and probabilistic-based methods, were developed, according to Chen and Yuan (1997), in response to the following

---

[2] [Oleson, 98], pp. xvi.

[3] Multiple technical disciplines involved in a common design endeavor. Expression used to reflect the interdisciplinary nature of complex systems design [Chen amd Lewis, 99].

31

concern: "How does one capture the uncertainty–which characterizes the early stages of design– and offers *flexibility* in specifying the design requirements so that the designs that are marginally outside the precise level of performance are not worthless?"

Chen and Lewis (1999) define their understanding of flexibility in the design process as follows:

> Our aim is to provide flexibility in the design process and to help further resolve the conflicts and disputes of rationality between the interests of multiple disciplines. By flexibility we mean that instead of looking for a single point solution in one discipline's model, we look for a range of solutions that involve information passing between multiple players (disciplines). With this flexibility, the design freedom of individual disciplines [...] could be significantly improved. Ultimately, this process will result in better products in less time because fewer iterations are needed.

Flexibility in the design process therefore entails expressing degrees of desirability in specifying design requirements. A recent report by the United States General Accounting Office on *Best Practices* in requirements specifications [GAO-01-288] echoes this description and emphasizes the need for a *flexible* behavior on behalf of the customers and developers in setting requirements:

> Flexibility in setting requirements is key to closing gaps between customer expectations and developer resources. While knowledge is essential to identifying gaps between expectations and resources, it takes *flexibility* on part of both the customer and the product developer to close the gaps. Flexibility represents the customer's ability and willingness to lower product expectations, coupled with the product developer's willingness and ability to invest more resources to reduce technical risks and other gaps before program start [...] In successful cases, requirements were flexible until the decision was made to commit to product development [...] This made it acceptable to reduce, eliminate, or defer some customer wants so that the product's requirements could be matched with the resources available to deliver the product within the desired cycle time.

2.2.3 Real Options and Managerial Flexibility

Today's market require that important investment decisions be made in very uncertain environments, when the market size, the time to market, the cost of development, the competitors' moves, and so on simply are not known.

32

Managerial flexibility[4] refers to the ability of management to affect the course of a project by acting in response to the resolution of market uncertainty over time. A flexible project may allow for downside protection against unfavorable market events, e.g., by abandoning the project, or introduce growth opportunities in the case of favorable conditions. Thus managerial flexibility reduces a project's exposure to uncertainty while providing management with the ability to respond to unfolding events. This concept is introduced in the context of Decision Tree Analysis and Real Options thinking. It is used in making a persuasive case against traditional valuation tools for capturing the value of staged or contingent investments (option to initiate a project, option to expand, to wait-and-see, etc.). A growing body of literature exists that describes the shortcoming of Discounted Cash Flow tools such as the classical NPV or IRR, and proposes ways of applying "Option Thinking" to valuing managerial flexibility. The reader is referred to Trigeorgis and Mason (1987), Triantis (1990), Faulkner (1996), or Amram and Kulatilaka (1999) for more elaborate discussions of option thinking and managerial flexibility.

On a parallel note, a plan of action is called rigid if it contains few contingent decisions, and flexible if it contains many such decisions. Plans made long in advance of the "action" are normally associated with rigidity, thus implying that to be flexible, one must be willing to wait and see, to defer decisions until one has taken into account the way a situation develops. Hence flexibility in this context implies remaining uncommitted to the extent of allowing oneself some leeway to design ways of dealing with unforeseen events [Rappaport, 1969].

2.3 Flexibility of a Design

A common theme across the previous discussions of flexibility is the ability to handle change. This characterization of flexibility however is not sufficient to distinguish it from other properties such as robustness. The ambiguity arises from the ill-defined term "change". A clear definition of flexibility should provide the following information:

- A time reference associated with the occurrence of change, i.e., when is the "change" happening during the life cycle of the system.
- A characterization of what is changing, e.g., the system's environment, the system itself, or the customer's needs of the system.

---

[4] The expression was first introduced by Trigeorgis and Mason in "Valuing Managerial Flexibility" in *Midland Corporate Finance Journal*, 5-1987. pp. 14-21.

- An indication for providing metrics of flexibility, or the ability to rank different designs according to their flexibility.

2.3.1 Time Frame Attached to a System's Life Cycle

A system's life cycle starts with the identification of customer's needs and proceeds towards the definition, design, production, operations, and disposal of a particular system. Prior to fielding, the process needn't be sequential: Different development models exist, e.g., the waterfall model, the spiral model, that offer a particular perspective, insights, and solutions to product development lifecycle problems. Each model generally constraints the sequence in which work is performed starting when the product is conceptualized and ending when the product has satisfied the acceptance criteria [Requirement Management Handbook, 96].

Needs identified	→	Conceptual design	Preliminary system design	Detailed design and definition	Production	System operations	System retirement
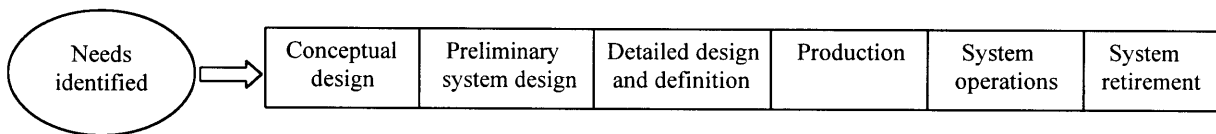
Fig. 2-2. Example of a system life cycle. Adapted from [Blanchard, 98]

In the particular case of a space system, the life cycle typically progresses through four phases [Wertz and Larson, 99]:

- Concept exploration, the initial study phase of a space mission which results in a broad definition of the space mission and its components.

- Detailed development, the formal design phase, which results in a detailed definition of the system components and, in larger programs, development of test hardware or software.

- Production and deployment, the construction of the ground and flight hardware and software and launch of the full constellation of satellites.

- Operations and support, the day-to-day operation of the space system, its maintenance and support, and finally its de-orbit or recovery at the end of the mission life.

These phases are named differently depending on whether the sponsor is NASA or DoD or some other agency:

Pre-Phase A Advanced studies	A Preliminary analysis	B Definition	C Design	D Development	E Operations

MCR   MDR   PDR   CDR   ORR   DR

Pre-Milestone 0	0 Concept exploration	I Demonstration and validation	II Manufacturing development	III Production and deployment	Operations and support

SRR   PDR   CDR
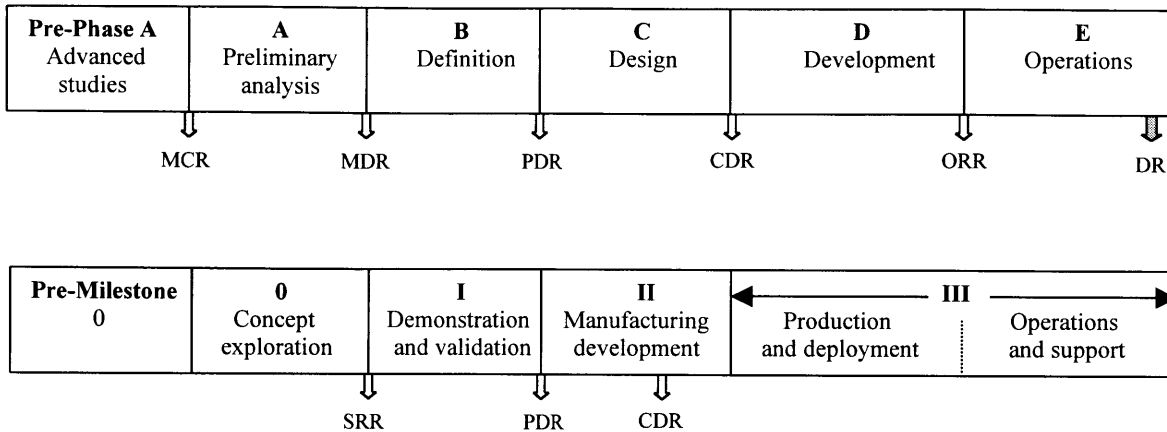
Fig. 2-3. Above, NASA's space program development phases and the associated "gates" or milestones of the program: MCR mission concept review, MDR mission design review, PDR preliminary design review, CDR critical design review, ORR operational readiness review, DR decommissioning review. Below, the DoD's development phases: SRR system requirement review. Not all the program's milestones are represented. Adapted from [Wertz and Larson, 99].

The system's life cycle provides an appropriate time reference for our purposes as described below:

Conceptual design	Preliminary system design	Detailed design and definition	Production	System operations	System retirement

T_0 T_{prod} T_{ops} Time

Mitigating risks of requirement changes occurring in this period is subsumed under process flexibility

Handling requirement changes occurring in this period is relevant to flexibility of a design
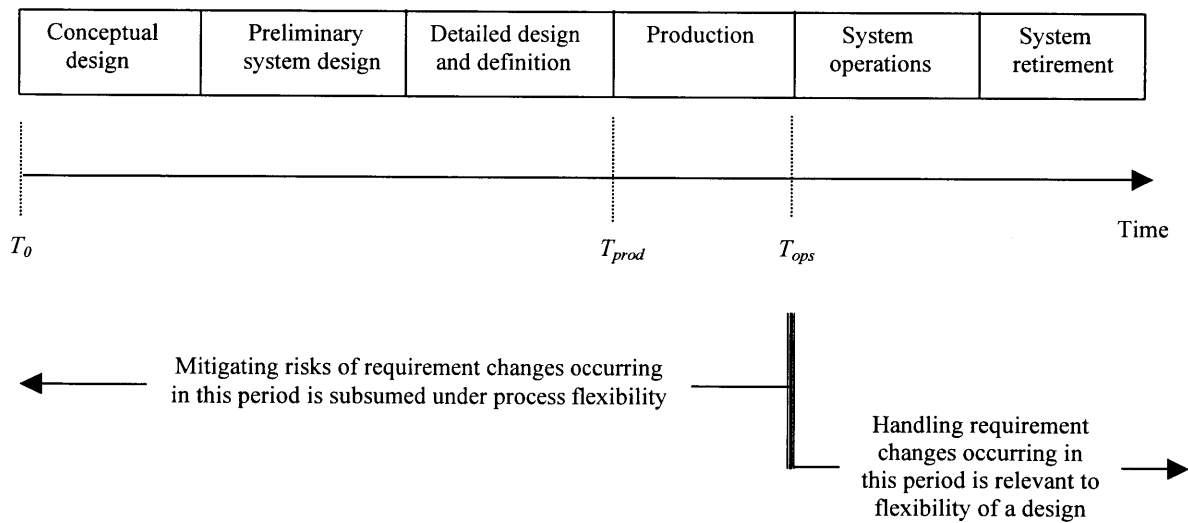
Fig. 2-4. Time frame *attached* to a system's life cycle, and time periods associated with process flexibility versus flexibility of a design.

35

Although the development process is rarely sequential, the program's milestones, e.g., the preliminary design review (PDR) and the critical design review (CDR), ensure that T_{prod} and T_{ops} are well defined in the case of space systems[5]. Changes may occur any time.

Current research that have addressed the issue of achieving flexibility in the multidisciplinary design [Chen and Yan, 99], [Lewsi and Mistree, 99], [Chen and Lewis, 99] have dealt with different ways of specifying requirements and handling their dynamics or changes occurring prior to T_{ops}. This was undertaken in order to resolve the conflicts of rationality between the interests of multiple disciplines involved in a common design endeavor. This is the time period—prior to T_{ops}—with which "flexibility in the design process" is concerned. Process flexibility include activities, methods and tools devised to mitigate the risks—cost, schedule, and performance—resulting from requirement changes occurring before fielding a system (see Fig. 2-4). This is further discussed in §3.2.

This is not the focus of this work. We will mainly be concerned with changes occurring after T_{ops}. But what are these changes about? Changes can occur in the system's environment (political, cultural, organizational, physical, etc.), in the system itself (e.g., wear and tear), or in its requirements—capabilities and attributes—resulting from changing customer needs.

2.3.2 Definition: Flexibility of a Design

We define flexibility of a design as the property of a system that allows it to respond to changes in its initial objectives and requirements—both in terms of capabilities and attributes—occurring after the system has been fielded, i.e., is in operation, in a timely and cost-effective way.

A discussion of "systems", "design", and "systems engineering" is provided in Appendix A. "Requirements", "capabilities" and "attributes" are used in the sense defined by the IEEE Standard 1233, 1998 Edition:

> A requirement is:
>
> (a) A condition or capability needed by a user to solve a problem or achieve an objective.

---

[5] In the case of other artifacts, T_{ops} is always well defined irrespective of the development model, e.g., waterfall, or spiral model.

36

(b) A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed document.

(c) A documented representation of a condition or capability as in definition of (a) or (b).

Requirements can be taken from customer needs and can be derived from technical analysis.

Capability (or functionality): Capabilities are the fundamental requirements of the system and represent the features or functions of the system needed or desired by the customer. A capability should usually be stated in such a way that it describes what the system must do. The capability should be stated in a way that is solution independent.

This definition provides a mean for distinguishing between requirements as capabilities and the attributes of those requirements. The following example is provided in the IEEE Guide for Developing System Requirements Specifications (1998) as a well-formed requirement:

For example:

Requirement: Move people from New York to California at a maximum speed of 5300 km/hr [IEEE Standard 1233, 98].

> *Capability*: Move people between California and New York
> *Attribute*: Cruising speed of 2500 km/hr
> *Constraint*: Maximum speed of 5300 km/hr

Requirement: The Mars Global Surveyor Spacecraft shall be capable of providing delta-V of 1290m/s, inclusive of finite burn losses from thrust vector misalignment, gravity losses, and all other maneuver inefficiencies [MGS Spacecraft Requirements, JPL D-11509].

> *Capability*: Provide propulsive capability
> *Attribute*: Delta-V = 1290m/s
> *Constraints*: Despite various losses (thrust vector misalignment, gravity, etc.)

A corollary of our definition of flexibility is that a flexible system can be modified in a timely and cost-effective way in order to satisfy different requirements at different points in time. These requirements, or requirement changes, as well as the time of occurrences of these changes, can be known or unknown a priori.

Examples of flexible designs will be discussed shortly after the concept of flexibility is disentangled from that of robustness and universality.

2.3.3 On Robustness: A Brief Survey of Robust Control and Robust Design

As stated previously, the distinction between the two concepts, robustness and flexibility, is a subject rich with ambiguity. Any attempt to define flexibility should address this issue. In order to discuss this concern, the following paragraphs review the concept of robustness as devised in two major areas of engineering undertaking, namely in feedback control systems—Robust Control— and Robust Design, also known as Taguchi's method. The purpose of this discussion is present a conceptual understanding of **robustness** of a design so that it forms a background against which the above definition of **flexibility** can be contrasted.

2.3.3.1 Robust Control

Controls engineers have developed a set of sophisticated mathematical tools to handle **disturbances** and **model uncertainty** in systems they wish to control. "The main ingredients of present day robust control theory were already present in the classical work of Bode[6]" in 1945. The following discussion addresses some of the key ideas underlying Robust Control. The reader interested in the subject can review the work by Francis (1987), Doyle et al. (1992), Ackermann (1993), Dahleh et al. (1995), or R. Sanchez-Pena et al. (1998).

The goal of Robust Control and the essence of robustness—from a system's control perspective— are clearly stated by Stefani et al. (1994):

> The ultimate goal of a control-system designer is to build a system that will work in the real environment. Since the real environment may change with time–components may age or their parameters may vary with temperature or other environmental conditions–or

---

[6] R. Sanchez-Pena, M. Sznaier. *Robust Systems: Theory and Applications.* J. Wiley. 1998.

the operating conditions may vary–load change, disturbances–the control system must be able to withstand these variations.

Assuming the environment does not change, the second fact of life is the issue of model uncertainty. A mathematical representation of a system often involves simplifying and sometimes wishful assumptions. Nonlinearities are either unknown, and hence unmodeled, or modeled and later ignored to simplify the analysis. Different components of systems—actuators, sensors, amplifiers, gears, belts—are sometimes modeled by constant gains, even though they may have dynamics and nonlinearities. Dynamic structures, e.g., aircrafts, satellites, missiles, have complicated dynamics in high frequencies, and these may initially be ignored. Since control systems are typically designed using much-simplified models of systems, they may not work on the real plant in real environments.

The particular property that a control system must possess in order for it to operate properly [ensure stability and achieve a set of pre-defined performance specifications] in realistic situations is called **robustness**.

The above are some of the key conceptual issues Robust Control deals with. Next we examine some of the fundamentals of the Robust Design methodology. While these two fields—Robust Control and Robust Design—have rarely interacted, they have nevertheless manipulated similar concepts and dealt with comparable problems at some level of abstraction, even though their tools and their specific domain of applicability differ. The purpose of the following discussion is to extract the essence of robustness in the particular field of Robust Design.

2.3.3.2 Robust Design

Robust Design is a design methodology developed in order to make a product's performance insensitive to raw material variation, manufacturing variability, and variations in the operating environment [Phadke, 89]. It was developed in the late 1950s by Genishi Tagushi [Tagushi, 87][7] and builds upon ideas from statistical experimental design.

---

[7] G. Taguchi. *Systems of Experimental Design.* English translation edited by D. Clausing. UNIPUB/Kraus International Publications, New York, 1987.

Robust products work well even when produced in real factories and used by real customers under real conditions of use. For instance when buying a car, a customer wants one that will start readily in northern Canada in the winter and not overheat in southern Arizona in the summer for example. In other words, he or she wants a car that is robust with respect to variations of use conditions. He or she also prefers a car that is as good at 50,000 miles as when new, that is robust against time and wear [Phadke, 89]. The sources of undesirable variation, also called noises in this framework, are the following:

- Variation in conditions of use
- Deterioration or variation with time and use
- Production or manufacturing variations.

These three types of noises cause degradation of performance or deviation away from ideal customer satisfaction. In this context, robustness is a characteristic of a system that minimizes these deviations, keeping performance economically close to ideal customer satisfaction [Clausing, 94]. The *ideal quality* a customer can receive is that every product delivers the target performance each time the product is used, under all intended operating conditions, and throughout its intended lifetime [Phadke, 89]. Put differently, robustness is a characteristic of a system whose performance is least sensitive to variations in operating environment, variation in raw material, thus allowing the use of low grade material and components, and variation in manufacturing, thus reducing labor and material cost for rework and scrap.

One of the goals of Robust Design is to exploit nonlinearities in the relation between a product quality characteristic and the various product parameters and noise factors in order to find a combination of product parameters values that gives the smallest variation in the value of the quality characteristic around the desired target value. This can be easily understood using the following mathematical formulation. Let $\mathbf{x} = (x_1, x_2, \ldots x_n)^T$ denote the noise factors and $\mathbf{z} = (z_1, z_2, \ldots z_j)^T$ the product parameters–called controlling factors–whose values can be set by the designer, then if the quality characteristic of the product is given by:

$$y = f(x, z) \tag{2.1}$$

The deviation Δy of the quality characteristic from the target value caused by small deviations Δx_i of the noise factors from their nominal values can be approximated by the first terms of the

Taylor series expansion of $f(x, z)$ around x_0 and z_0 where x_0 is the expected value of the noise factors and z_0 the unknown nominal settings of the product parameters:

$$\Delta y \cong \left(\frac{\partial f}{\partial x_1}\right)\Delta x_1 + \left(\frac{\partial f}{\partial x_2}\right)\Delta x_2 + ... + \left(\frac{\partial f}{\partial x_n}\right)\Delta x_n \qquad (2.2)$$

The partial derivatives are evaluated at x_0 and z_0. The above notation for the sensitivity coefficient is shorthand for:

$$\left(\frac{\partial f}{\partial x_i}\right) \equiv \left(\frac{\partial f}{\partial x_i}\right)_{x_0, z_0} \qquad (2.3)$$

If the deviations in the noise factors are uncorrelated, the variance of quality characteristic can be expressed in terms of the variances of the individual noise factors as follows:

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1}\right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2 \sigma_{x_2}^2 + ... + \left(\frac{\partial f}{\partial x_n}\right)^2 \sigma_{x_n}^2 \qquad (2.4)$$

Thus the variance of the quality characteristic is the sum of the products of the variances of the noise factors times the sensitivity coefficients. The sensitivity coefficients are themselves function of the control factors as expressed in Equation (2.3). As can be seen from (2.4), the variance of the quality characteristic can be minimized by either selecting the control factors z_0 such that the sensitivity coefficients are minimum, or by reducing the variances of some of the noise factors, typically the tolerances on system's components. The first action is referred to as **parameter design**; the second is called **tolerance design**. Figure 2-5 illustrates the difference between achieving robustness via parameter design versus tolerance design.

41
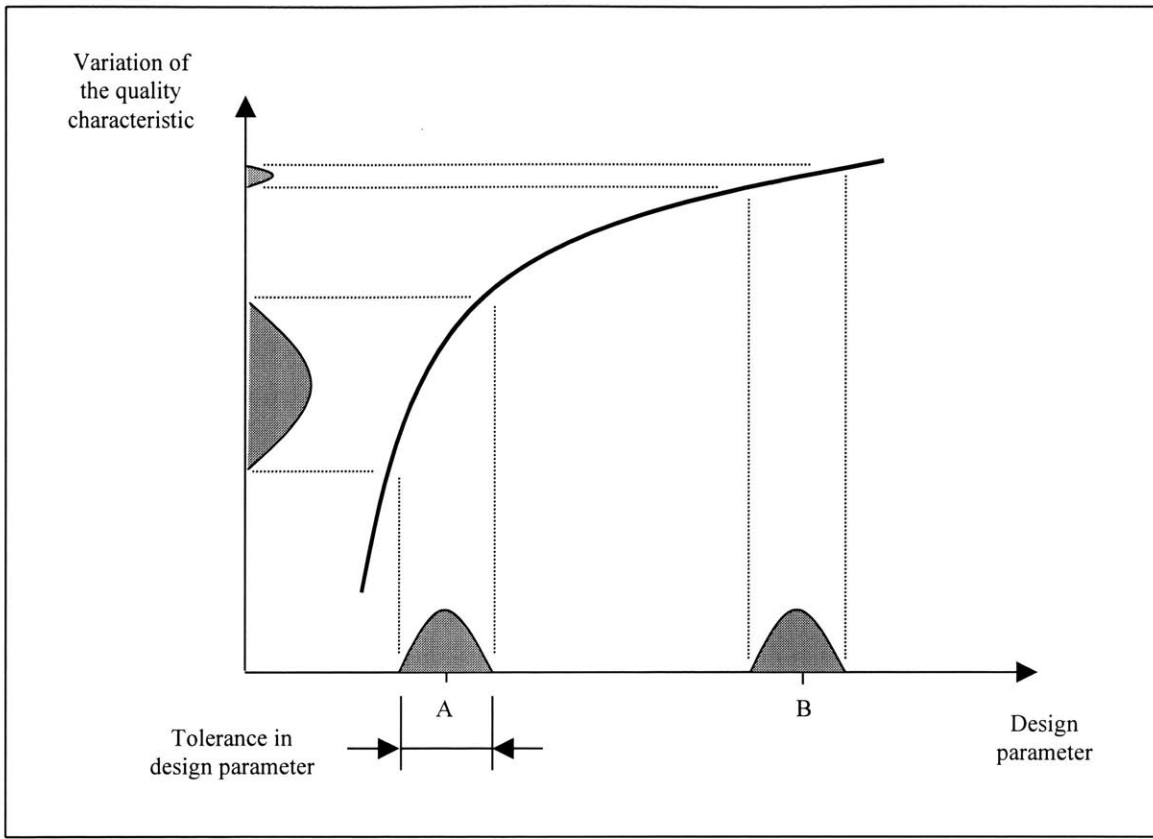
Fig. 2-5. Variation of the quality characteristic as a function of the design parameter setting.

The above plot shows that a reduction in the variation of the quality characteristic of a product can be achieved by appropriately setting the design parameter, or control factor, at a point where the sensitivity—defined in Eq. 2-3—is small, e.g., choosing point B over point A. This is referred to as parameter design and does not affect the manufacturing cost of the component, as opposed to tolerance design that consists of reducing the tolerances on the design parameters, and is associated with more costly parts. From this perspective, it is clear that parameter design should be carried out prior to tolerance design in order to deliver robust products. This is a fundamental idea in Robust Design.

2.3.3.3 Synthesizing a Definition of Robustness

From the previous discussion, we can synthesize a general definition of robustness as the property of a system which allows it to satisfy a **fixed** set of requirements, **despite** changes occurring after

the system has entered service, in the environment or within the system itself, from the nominal or expected environment or the system design parameters[6].

For instance, in the case of Robust Design, the objective is to maintain a target performance despite the various noise factors such as the variations in the conditions of use of the system, the degradation of the system or system components with time, and the manufacturing variability. In the case of Robust Control, the fixed set of requirements is ensuring stability and maintaining some pre-defined performance specifications.

2.3.4 Distinction Between Flexibility and Robustness of a Design

The definitions discussed above provide a clear distinction between robustness and flexibility of a design. Although these two concepts refer to the ability of a system to handle change, the nature of the change, as well as the system's *reaction* to the change, in each case is very different: Flexibility, as defined herein, implies the ability of a design to satisfy **changing requirements** after the system has been fielded, whereas robustness involves satisfying a **fixed set of requirements** despite changes in the system's environment or within the system itself. The relation between flexibility and robustness of a design as a function of the system's objectives and environment is graphically illustrated in Figure 2-6.

System's objectives after fielding

	Fixed/known	Changing and/or unknown
Changing	Poor Design!	Flexible Design
Fixed	Optimized Design	Robust Design
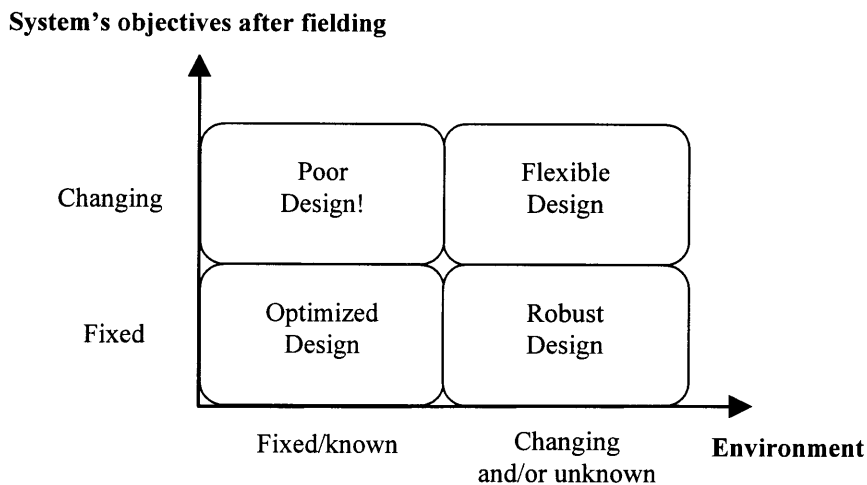
Environment

Fig. 2-6. Flexibility and Robustness as a function of the system's objectives and environment.

---

[6] Design parameters are defined in [Suh, 99] as the key physical variables that characterize a design and satisfy a set of specified requirements.

43

The following thought experiment would help clarify the distinction between flexibility and robustness of a design. Imagine designing a spacecraft for 50-100 years! The two major challenges in striving for such a spacecraft design lifetime are the following:

1. Maintain on-board functionalities after launch, despite changes in software and hardware characteristics due to radiation impacts, malfunctions, aging, etc. This is indicative of the need for robustness of the design, i.e., robustness has to be built-in into the spacecraft.

2. Create new functionalities on-board for changes in requirements occurring after launch, as events unfold, new environments are explored, and/or new data becomes available, etc. Such changes are bound to happen given the extensive spacecraft design lifetime. This is indicative of the need for flexibility of the design, i.e., flexibility has to be embedded in the spacecraft.

2.3.5 Distinction Between Universality and Flexibility of a Design

Another distinction can be made, based on the definition of flexibility provided above, between two concepts that are potentially to be confused with one another: That of flexibility versus universality of a design. Software for instance that can be used in a variety of situations **without change or modification**, is considered "universal" not flexible. *Flexible* software [Parnas, 78] is one that can be easily changed—extended, contracted, or else—in order to be used in a variety of ways. Similarly, spacecraft that carry multiple instruments and perform multiple missions simultaneously are NOT considered flexible according to the definition of flexibility provided above. Likewise, a design is considered flexible if it is **easily changeable** to be used in a variety of ways. The time and cost required to implement the changes are two indicators of the "ease of change" of a design and reflect its flexibility.

2.4 Examples: Flexibility and Product Design Lifetime

The following examples illustrate the relationship between flexibility and design lifetime. The first example contrasts the operational lifetime of the Boeing B-52 with that of the Convair B-58, and makes the case that the B-52 was a highly flexible design[7]. The second example discusses the need for flexibility in the rotorcraft industry, particularly in the light of the current restrictive military spending and the fact that helicopters are being designed for increasingly longer

---

[7] Although perhaps its flexibility was accidental.

44

lifetimes. The third example argues that the Galileo spacecraft, by completing its initial objectives and performing a new or extended mission, constitutes an instance of flexibility in space systems.

2.4.1 Designing for Flexibility: The Boeing B-52 versus the Convair B-58

In order to illustrate the relationship between a product's life-span, the initial circumstances from which the system's requirements were derived and the various environments in which it can operate, the Boeing B-52 Stratofortress is presented as an example of a flexible design, and is contrasted with the Convair B-58 Hustler. The purpose of this section is not to delve into the particular design practices that enabled the B-52 to remain in operation long after the B-58 was retired, but simply to illustrate the above-mentioned relationship.

The B-52 is a long-range, heavy bomber that can perform a variety of missions. It is capable of flying at high subsonic speeds (Mach 0.86) at altitudes up to 50,000ft, and carry both conventional and nuclear ordnance. In a conventional conflict, the B-52 can perform a variety of missions such as air interdiction, offensive counter-air, or maritime operations. It is capable of dropping or launching the widest array of weapons in the U.S. inventory including gravity bombs, cluster bombs, and guided missiles [Boeing B-52 Stratofortress, 96]. The venerable aircraft has also been used to ferry both manned and unmanned systems for altitude drop and orbital insertion.

The B-52 first entered service in 1955 with the Strategic Air Command. The initial specifications were issued on November 23rd 1945. For the first 10 years of its Air Force service, it operated in a cold war atmosphere. Current engineering analysis shows the B-52 life-span can be extended beyond the year 2045. Thus it will be a "century" aircraft. It has assumed important conventional roles in Vietnam and the Gulf war. These are very different environments from which the initial system requirements were derived—different environments thus different threats, hence the need to alter the tactics in order survive and prevail—No other weapons system offers the flexibility of the B-52. It is referred to as the bomber that "is not getting older, just getting better" because it was capable of accommodating numerous improvements over the years. Upgrades since the early 1980's have included many new and improved systems:

- Offensive avionics
- Environmental control
- Auto-pilot

- Enhanced electronic countermeasures
- Conventional air-launched cruise missile (CALCM)

The Convair B-58 Hustler on the other hand was the first supersonic bomber to enter service with the USAF in March 1960. Despite its high performance and sophisticated equipment, the service of the B-58 was brief; the aircraft flew for only a decade before being consigned to storage. Part of the reason for this rather short service was due to the aircraft's rather high accident rate. Another factor was the intercontinental ballistic missile, which entered service at the same time as the B-58 and removed its primary mission. Of course the same was true of the B-52 but it proved flexible enough to find widespread use in other mission areas. Aside from the technical problems that plagued the B-58, the aircraft in some sense lacked the flexibility of the B-52 to adapt to new missions and roles in new environments.

It is tempting at this point to probe the original requirements of both the B-52 and B-58, and to identify the particular design choices that rendered on one hand the B-52 a flexible design to remain in operation for almost a century, and on the other hand the B-58 a short-lived inflexible design. The study should investigate for example the impact of the requirement to fly at supersonic speeds for the B-58 on the wing design and the airframe, and how this choice, later during the operational life of the B-58, prevented it from accommodating different weapons and performing other missions than the one the it was initially designed for. This is however beyond the scope of this section. The purpose of this example, as stated above, is to illustrate the relationship between flexibility and product life-span, and not to delve into the particular design choices that render a product flexible[8].

2.4.2 The Need for Flexibility in the Rotorcraft Industry

Helicopters tend to have an operational life-span exceeding 30 years. In many cases, this is long after the circumstances for the original requirements have been removed. Hence, the missions and roles of a rotorcraft are most likely to change over its life-span. Also, the embedded technologies within the rotorcraft continue to evolve after the product has been fielded. Due to the high value of these already fielded products, there is a tendency among the operators to modify the fielded rotorcraft to adapt to new missions and roles as opposed procuring new ones. Furthermore, the

---

[8] It is also possible that the complete story of the short-lived B-58 may not yet be known: An anonymous reviewer pointed out that the B-58 was consigned to storage for classified reasons.

46

current state military spending in the United States forces traditional military contractors to seek non-traditional business segments of the market. Therefore fewer products have to be designed for an extended life-span and with the ability to perform new and diverse missions. Rotorcrafts in particular have to be designed with the ability to be modified after entering service in order to perform new and emergent missions. In other words, flexibility has to be embedded in the initial design.

Consider for instance the Sikorsky medium lift helicopter S-70 or its UH-60 designation for its military role. The helicopter was developed in the early 1970s in response to the U.S. Army rotary-winged aircraft program referred to as the Utility Tactical Transport Aircraft System (UTTAS). In short, the UTTAS program required a helicopter to perform multiple missions such as troop transport, air cavalry, and medical evacuation. The program also included standards for the helicopter's combat survivability, reliability, maintainability, as well as adverse weather and nighttime operational capabilities. Designed as a military utility helicopter, the UH-60/S-70 has now over 35 derivatives performing a variety of missions, e.g., troop transport, cargo movement, medical evacuation, VIP transport, and has been sold in over 90 countries [Holmes, 99]. The need to access new markets, or to satisfy specific customers requirements, led to the development of these derivatives. However, it is the intrinsic ability of the UH-60/S-70 baseline architecture to accommodate changes following new customers requirements—in a timely and cost-effective way in order to achieve a different configuration vehicle—that made it possible to develop these derivatives. Holmes (1999) argues that the use of platform design for medium lift helicopters enabled the expansion of the mission roles and capabilities of the UH-60/S-70 thus provided the flexibility of its baseline architecture.

2.4.3 Galileo's Mission to Jupiter and the Galileo Europa Mission (GEM) Extension

The Galileo spacecraft is a NASA robotic mission to explore Jupiter. The spacecraft consisted of an orbiter and an atmospheric entry probe designed to enter Jupiter's atmosphere and provide a weather report on temperature, pressure, composition, wind, and lightning of Jupiter's atmosphere. The spacecraft was launched on-board the Space Shuttle Atlantis in 1989 and reached Jupiter in 1995.

The initial science objectives of the Galileo orbiter included the following:

    1a. Investigating the circulation and dynamics of the Jovian atmosphere and ionosphere

2a. Characterizing the vector magnetic field and the energy spectra, composition, and distribution of energetic particles and plasma to a distance of $150R_J$

3a. Conducting long-term observation of its magnetosphere

4a. Characterizing the morphology, geology, and physical state of the Galilean satellites [Galileo Project Information, 99][9]

On December 1997, Galileo successfully completed its original mission objectives: A two-year study of the Jovian system. Since the resilient spacecraft was capable of much more, it was decided to extent the mission, now called the Galileo-Europa Mission (GEM), in order to study in detail Jupiter's icy moon Europa and its fiery moon Io. The new major science objectives of the GEM are the following:

1b. Europa: Study and characterize crust, atmosphere, and possible ocean (i.e., implication for exobiology) using imaging, gravity, and space physics data

2b. Io Plasma Taurus: Explore and map Io Plasma Taurus as orbit approaches Io.

3b. Io: Intensive study of Io's volcanic processes, atmosphere, and magnetosphere environment [GEM Fact Sheet, 00][10]

The fact that the orbiter has completed its initial mission and performed a *new* or extended mission constitutes one instance of flexibility of a space system as defined above—ability to respond to changes in a system's initial objectives and requirements occurring after the system has been fielded. This flexibility was in part due to the various design margins that the orbiter had, e.g., its design lifetime exceeded the time required to complete its science objectives (ΔV margin, etc.). Other instances of flexibility in space systems are discussed below.

2.5 Flexibility in the Context of Distributed Satellite Systems (DSS) and TechSat21

Distributed space architectures, or the spreading of functionalities across multiple spacecraft, thus forming a virtual satellite, enable new missions to be performed, and often offer reduced cost or improved capabilities over monolithic designs. Martin and Stallard (1999) discuss the application of DSS to synthesize a large space aperture:

---

[9] Can be found at www.nssdc.gsfc.nasa.gov/planetary/galileo.html

[10] Can be found at www.jpl.nasa.gov/galileo/gem/fact.html

Since the satellites are not connected by structures, they can be separated over very large baselines that could not be considered for monolithic apertures. This feature can be beneficial for such missions as space-based radar, or large apertures for detection of slow moving targets in clutter. [...] Another mission application [of DSS] is mobile jam resistant communications [...] or interferometric imaging..."

The ability to reconfigure a cluster's geometry for instance allows modifying the revisit time requirement. This in one particular instance of flexibility—ability to respond to changes in the requirements occurring after the system has been fielded—that is characteristic of DSS and that is not feasible with a monolithic design. Furthermore, the ability to modify the revisit time (RT) on-orbit implies that it needn't be specified prior to launch or further up-front in the development phase of the system.

The idea that critical system requirements need not be narrowly specified prior to launch, because changes can be accommodated afterwards, is one particular advantage of the property of flexibility in design. It seems particularly important and valuable in defense oriented space systems for instance where the development times are of the order of 5 to 10 years, and changes are very likely to occur, as well as for systems that operate in uncertain environments.

TechSat21 is an Air Force Research Laboratory program designed to explore new technologies for lightweight and low-cost clusters of micro-satellites. One instance of flexibility of TechSat21 for example results from the ability to modify of the cluster geometry in order to operate in a Geo-location mode instead of the nominal Radar mode. This is illustrated in the figure 2-7.

Radar Mode   Geo-location Mode
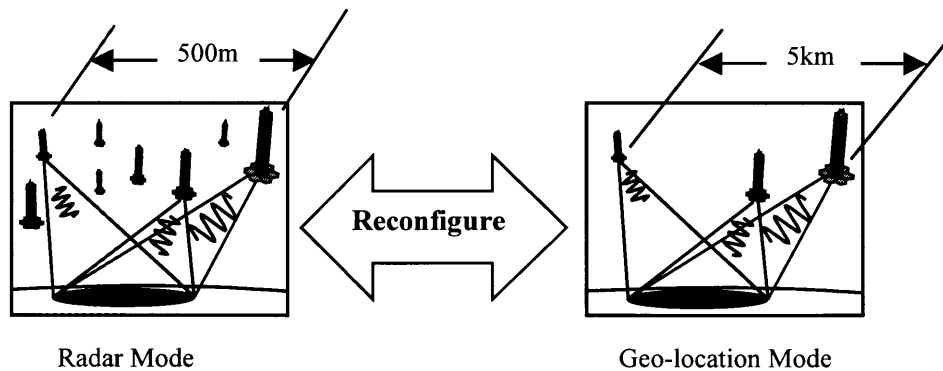
Fig. 2-7. Reconfiguring the cluster geometry allows "other" missions to be performed. This is not feasible with a monolithic space system design.

49

Figure 2-8 illustrates two different types of flexibility associated with TechSat21: The first involves the ability of the system to change its mode of operation—in the parlance of the IEEE Standard 1233, this is relevant of the system's **capability**—whereas the second type involves the ability to modify the **attribute** of the requirement (tune-in performance).
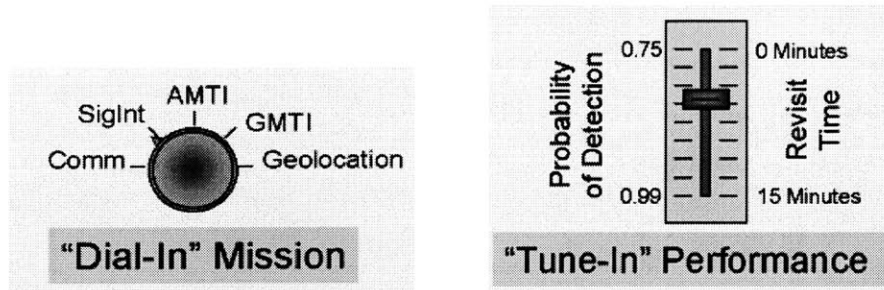
Fig. 2-8. Two generic types of flexibility: The ability to modify the mode of operation of a system (Dial-In Mission), and the ability to modify the attribute of a requirement (Tune-In Performance).

2.6 Summary and Conclusions

This chapter reviewed the concept of flexibility as discussed in various fields of investigations and extracted its characteristics features. Flexibility of a design is here defined as the property of a system that allows it to respond to changes in its initial objectives and requirements—both in terms of capabilities and attributes—occurring after the system has been fielded, i.e., is in operation, in a timely and cost-effective way.

In order to discuss any subject matter clearly, it is necessary to begin with a clear set of definitions. Indeed much can be gained through careful and consistent definitions of terms alone. Flexibility however has been a word rich with ambiguity. The first section of this chapter identified the various sources of ambiguity in discussions of issues of flexibility: These include the failure of the familiar characterization of flexibility—ability to handle change—to distinguish it from other properties, particularly in the light of the proliferation of its pseudo-synonyms. A selected literature review is then provided and a definition of flexibility is synthesized.

A brief literature review of Robust Control and robust Design is also presented. Robustness is defined as the property of a system that allows it to satisfy a **fixed** set of requirements, **despite** changes occurring after the system has entered service, in the environment or within the system itself, from the nominal or expected environment or the system design parameters.

Robustness and flexibility are then contrasted, and a distinction is drawn as well between flexibility and universality of a design. Flexibility of a design is also disentangled from issues of flexibility in the design process: The latter include activities, methods, and tools devised to mitigate the risks—cost, schedule, and performance—resulting from requirement changes occurring during the design process, i.e., before fielding a system.

Several examples of flexible systems are finally discussed, and illustrate the relationship between flexibility and a system's design lifetime. The examples included the (accidentally) flexible B-52—to remain in operation for almost a century—versus the short-lived inflexible B-58, the Galileo spacecraft; its initial mission to Jupiter and its extended mission to Io and Europa (the Galileo Europa Mission), as well as instances of flexibility in the context of distributed satellite systems.

Aside from the particular points referred to above, this chapter also laid a framework for a clear and comprehensive discussion of issues of flexibility: One may disagree with the particular definitions provided herein, that's fair. The reader is encouraged to create his or her own set of definitions but should make sure that they are unambiguous, self-consistent, and lead to useful concepts. However a comprehensive treatment of flexibility in system design should address the following questions:

1. What is flexibility?
2. Why or when is flexibility needed in system design?
3. How can we design for flexibility? What are the design principles for embedding flexibility in system design?
4. What are the trade-offs associated with designing for flexibility? What is the value of flexibility and what are the penalties (cost, performance, risk, etc.), if any, associated with it?

This chapter focused on the first question and extracted the characteristics features of flexibility. The following chapter will address the second question: It will make the case for flexibility in system design by identifying and characterizing the situations in which flexibility in system design is needed.

Chapter 3

The Case for Flexibility in System Design

"It has always seemed to me that this northern route to the City of Mexico would have been the better one to take.
But my later experience taught me two lessons:
First, things are seen plainer after the events have occurred;
Second, that the most confident critics are generally those who know the least about the matter criticized."

General U. S. Grant on the Mexican War, *Personal Memoirs*, 1885.

The previous chapter reviewed the concept of flexibility as discussed in various fields and proposed a definition of flexibility in system design. A distinction was drawn between process flexibility and flexibility of a design. This definition of flexibility was then used to disentangle the concept of flexibility from *robustness*, as well as from *universality*. As formerly discussed, a comprehensive treatment of flexibility in system design should tackle the following questions: 1) What is flexibility? 2) Why or when is flexibility needed in system design? 3) How can one design for flexibility or what are the design principles for embedding flexibility in system design? 4) Finally, what are the trade-offs associated with designing for flexibility, i.e., what is the value of flexibility and what are the penalties (cost, performance, risk, etc.), if any, associated with designing for flexibility? This chapter addresses the second question by identifying and characterizing the situations in which flexibility in system design is needed.

3.1 Introduction: The Need for Flexibility in High Value On-Orbit Assets

In recent years, several space programs have chosen to increase their space segment design lifetime. Over the last two decades, communication satellites in the geo-stationary ring for instance have seen their design lifetime on average increase from seven to fifteen years. This

trend is also observed in the design and development of many high value assets, e.g., the average life span of a helicopter delivered today can exceed thirty years or 20,000 hours of operation.

In most cases, increasing the space segment design lifetime was driven by–and justified by– traditional economic analysis using Discounted Cash Flow techniques such as the classical Net Present Value (NPV), and the desire to maximize the Return On Investment (ROI). However, extending the design lifetime has several side effects. On one hand, it leads to larger and heavier satellites resulting from several factors such as additional propellant for orbit and station-keeping, excess power generation and storage, etc. This in turn increases the satellite's development and production costs. On the other hand, satellites with long design lifetimes can become obsolete, technically and commercially, before the end of their mission (see Fig. 3-1). In many cases, the initial circumstances from which the original system requirements were derived change or are modified during the spacecraft's operational lifetime.
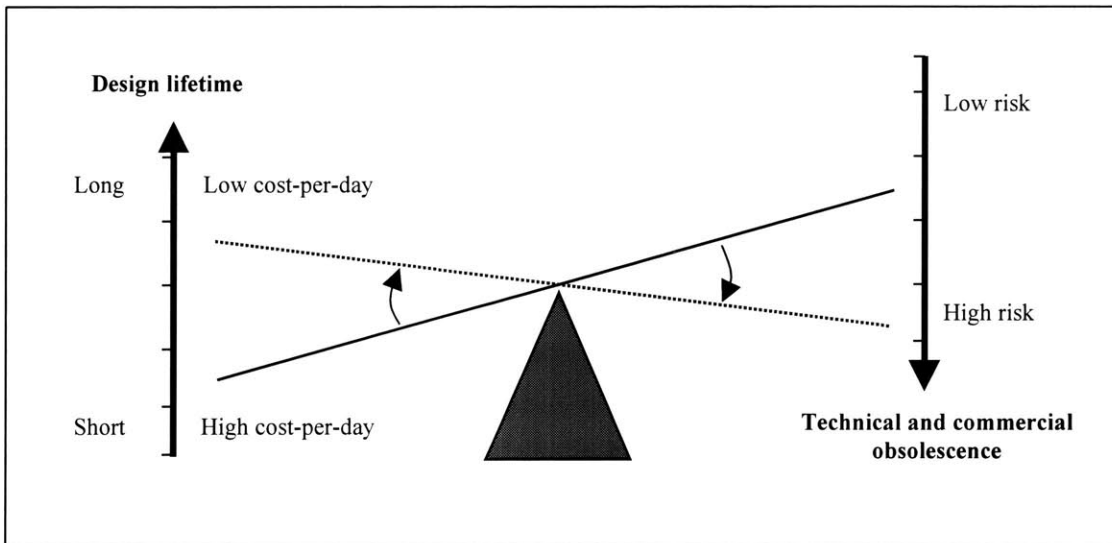
Fig. 3-1. Graphical illustration of the design lifetime trade-off: Extending a satellite design lifetime decreases its cost per operational day. However, it increases the risk that the satellite becomes technically and commercially obsolete before the end of its lifetime.

Due to the high value of these on-orbit assets, it is desirable to have space systems that are flexible and can adapt to new or emergent missions and roles, instead of launching new ones. Flexibility is thus a key property we seek to embed in high-value on-orbit assets. This example points out the guidelines to make the case for flexibility in system design. Namely, three issues, as well as the relationship between them, have to be investigated:

53

- The product design lifetime
- The dynamics of the environment the system is operating in–on a time scale comparable to that of the product design lifetime–as well as the uncertainty that plagues this environment. For instance, in the case of a commercial venture, this refers to the dynamics of the market the system is serving as well as the uncertainty or volatility that characterizes this market.
- The dynamics of the product's technology base (rate of change of technology, time to technology obsolescence, etc.)

In the following, it will argued that the need for flexibility in system design, as previously defined, arises for systems that have to operate in highly uncertain environments and/or when the system's design lifetime considerably exceeds the time constants associated with the market dynamics the system is serving in the case of a commercial system and the various technologies embedded in the system. More generally, it will be argued that flexibility should be sought when:

1. The uncertainty in a system's environment is such that there is a need to mitigate market risks, in the case of a commercial venture, and reduce a design's exposure to uncertainty in its environment[1].

2. The system's technology base evolves on a time scale considerably shorter than the system's design lifetime, thus requiring a solution for mitigating risks associated with technology obsolescence.

In other words, **flexibility reduces a design's exposure to uncertainty, and provides a solution for mitigating market risks as well as risks associated with technology obsolescence.**

This chapter is organized as follows: Section 2 recalls the distinction between process flexibility and flexibility of a design, and discusses the dynamics of system requirements during the requirement generation phase. This in turn is used to make a brief statement on the need for process flexibility. Section 3 addresses the need for flexibility in system design. First the case of

---

[1] This is an all-encompassing statement and includes a wide range of activities and rationales. "Environment" is used in the sense defined by the IEEE Std 1233 and consists of "circumstances, objects, and conditions that will influence the completed system. They include political, market, cultural, organizational, and physical influences as well as standards and policies that govern what the system must do or how it must do it." [IEEE, 98a]

flexibility is made in relation to fast market dynamics (on a time scale of the order of the system design lifetime); the failure of the Iridium system is discussed as an example of the consequences of the lack of flexibility of a system designed to operate in a rapidly evolving market. Second, it is argued that flexibility should be sought when the system's technology base evolves on a time scale considerably shorter than the system's design lifetime. A discussion of technology evolution and obsolescence is provided, and a metric is introduced to quantify the disparity between components' life cycle (or Time to Obsolescence) and the system's design lifetime, thus driving the need for flexibility. Section 4 touches on the DoD's exposure to technology obsolescence problems. Section 5 contains the summary and conclusions.

3.2 Dynamics of System Requirements: Process Flexibility Versus Flexibility of a Design

Both process flexibility and flexibility of a design include the ability to handle changes in requirements. In order to understand what drives such changes, and when changes are triggered, it is necessary to first understand the process of generating requirements. A brief summary of this critical activity in system design is provided below.

3.2.1 Requirement Generation: A Critical System Design Activity

All requirements begin with succinct but well-defined user and customer needs [Wertz and Larson, 99]. The requirement generation phase is rightly perceived as a critical activity in the process of engineering a system. General practice is that requirements are defined, reviewed, and approved prior to system design.

Inadequate, ambiguous, and *unstable* system requirements are a major and continuing source of problems in systems development. These problems are manifest in missed schedules, budgets overruns, and systems that are to varying degrees, unresponsive to the true needs of the customer. These manifestations are often attributable to poorly defined or ill-understood processes used to elicit, specify, analyze, manage, and validate requirements [Requirement Handbook Guide, 96].

The requirement generation phase in a project is one of the most influential development steps with regard to eventual success of a program. It is known for instance that over 70% of the life cycle costs of a program are locked in before 10% is actually spent. Furthermore, the early stages of development offer the designer the most control over the eventual cost of the system. In addition, requirements generation has been found to be the development phase where many

55

programmatic problems originate (over 60% of systems error over a system life cycle). Two distinct characteristics of the requirement generation phase make it a critical area to research and improve because of the large payoff that is possible: Not only is the majority of the system life cycle cost committed during this phase, but also a great portion of the system errors can be traced back to it [Walton, 99].

The requirement generation phase is defined as "the formulation of requirements and consists of identifying the needs of the client and translating those needs into constraints, control, and measures for implementation" [White and O'Hair, 96 *from* Walton 99]. The requirement generation phase is a generic term and includes the following:

- Requirement Acquisition: This is the initial stage of requirement generation. It is the first opportunity to acquire customer needs, either by being given them, eliciting them, researching for them, or having a prior knowledge of them and understanding the intended use or needs. Some of the issues that arise in this phase include the identification of the customers and prioritizing the customer needs that are obtained.

- Requirement Analysis and Derivation: This consists of reviewing an existing set or requirements, deriving others, and allocating these requirements to the functional elements of a system.

- Requirement Correlation: The correlation of requirements involves the archiving of requirement information as well as understanding the relationships between them. Knowing where and when a requirement originated as well as its rationale and its relation to a customer need is an important discipline during the development phase.

- Requirement management: This part of the process involves collecting, documenting, and disseminating requirements. Managing requirements is necessary to ensure that no requirements are added or changed without proper authority. The process of managing requirements should be defined prior to the initiation of the project and should include who is responsible for collecting, archiving, and disseminating the requirements, as well as the review process for the acceptance or rejections of requirements changes.

3.2.2 Requirements Are Rarely Static

The requirement generation process interfaces with three entities: The user and the customer, the technical community, and the environment[2]. A simple graphical representation of the interactions between the various entities is shown in Figure 3-2.
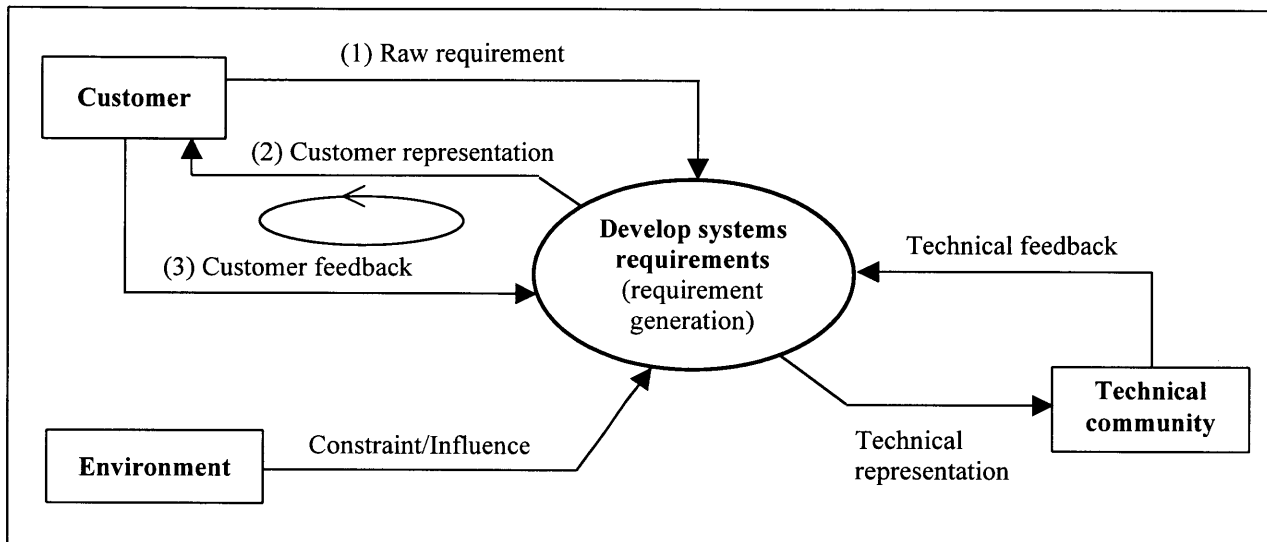
Fig. 3-2. Context for developing a System Requirements Specifications. Adapted from [IEEE, 98a]

The advantage of the simplicity of Figure 3-2 is that it illustrates where the "change", discussed in §3.2, can stem from, i.e., the customer, the environment, and/or the technical community. It is precisely this observation that requirements are rarely static that drives the need for both process flexibility and flexibility of a design as will be detailed further.

The IEEE Standard 1233 recommends freezing a system's set of requirements permanently early on in the life-cycle, but admits that this is rarely possible:

> Although it is desirable to freeze a set of requirements permanently, it is rarely possible. Requirements that are likely to evolve should be identified and communicated to both customers and the technical community. A core subset of requirements may be frozen

---

[2] The environment includes the circumstances, objects, and conditions that will influence the completed system. They include political, market, cultural, organizational, and physical influences as well as standards and policies that govern what the system must do or how it must do it [IEEE, 98a].

early. The impact of proposed new requirements must be evaluated to ensure that the initial intent of the requirements baseline is maintained [IEEE, 98a].

Figure 3-3 illustrates one rationale for fixing requirements early in the design process by representing the cost of a design change as a function of the time when the change was requested during the system's life cycle.
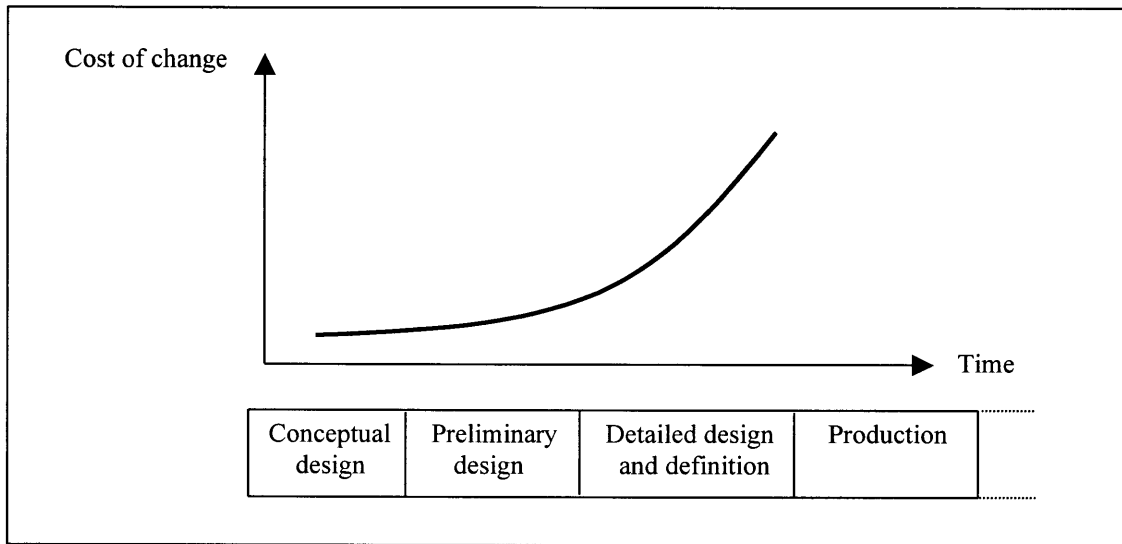
Fig. 3-3. The cost of an iso-change when introduced at different phases of a system's life cycle. Adapted from Blanchard (1998).

A similar plot is likely to result if one considers the effect of a change requested at different points during the system's life cycle on the development schedule. This illustrates why changes or instability in requirements are viewed as undesirable (negative impact on life cycle cost and development schedule). The standard approach to mitigate those risks has been to freeze the requirements early in the design process. The research discussed in §2.2.2 investigates other ways of coping with requirement changes occurring before fielding the product. Mitigating risks–cost, schedule, and performance–resulting from requirements changes occurring before fielding a system is subsumed under **process flexibility**. The ability to handle requirement changes occurring after a system has been fielded requires different techniques and reflects the **flexibility of a design**. The time periods associated with process flexibility versus flexibility of a design are illustrated in Figure 3-4 (see §2.3.1).

58

Conceptual design	Preliminary system design	Detailed design and definition	Production	System operations	System retirement

T_0 T_{prod} T_{ops} Time

← Mitigating risks of requirement changes occurring in this period is subsumed under process flexibility

Handling requirement changes occurring in this period is relevant of flexibility of a design →
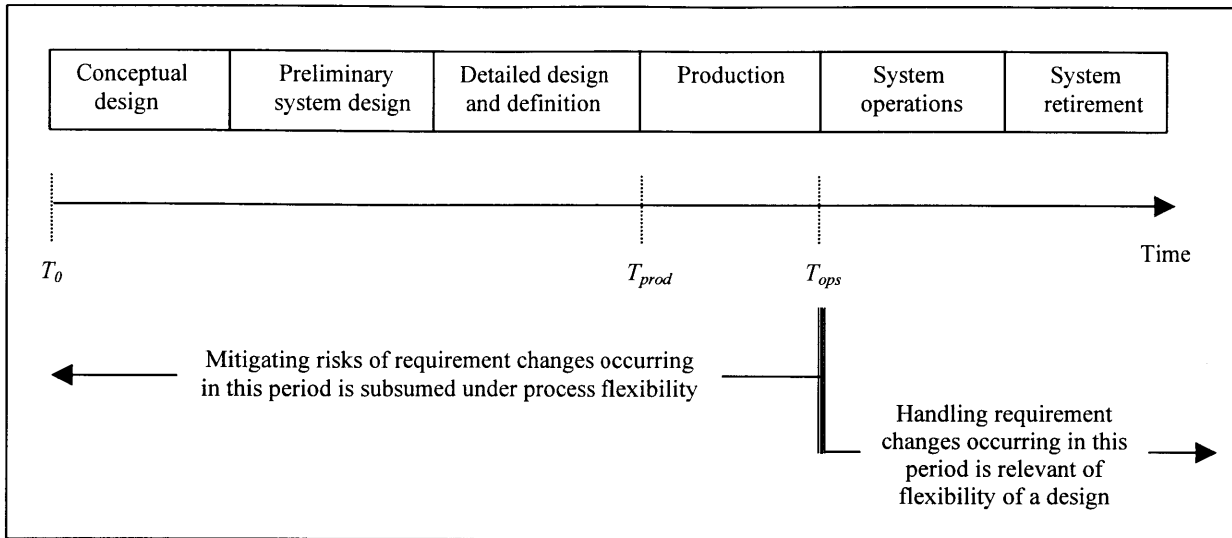
Fig. 3-4. Time periods associated with process flexibility versus flexibility of a design.

3.2.3 The Need for Process Flexibility

Although the need for process flexibility is not the focus of this chapter, it is nevertheless worth discussing even though briefly. Process flexibility includes activities, methods, and tools devised to mitigate the risks—cost, schedule, and performance—resulting from requirements changes occurring before fielding a system. The need for process flexibility therefore results from the need to handle requirement changes occurring **during the development phase**. This is particularly important in defense-oriented systems for example where development times of 10 to 15 years are frequent. Consider for instance the Army's Crusader artillery vehicle program, a self-propelled 155-millimiter howitzer and re-supply vehicle; the development program began in 1994 and production is expected to start in 2008 (14 years in development). Or the Comanche helicopter program for example, a lightweight, twin engine, stealthy helicopter; the development program was launched in 1988, and production is expected to start in 2006 [GAO-01-288].

Requirement changes are bound to occur during such long development phases because changes in the economic, strategic or tactical environment, as well as technological evolution will drive changes in customer's needs (see Fig. 3-2).

On the latter point (technology obsolescence and the need for process flexibility), as the time from design to production keeps stretching out, many of the technologies selected during the design are already obsolete before production starts. In the case of the F-22 for example,

59

production started more than **a decade after many of the design decisions were made** and components selected [Hitt and Schmidt, 98].

Given that changes in customer's needs are unavoidable during the development phases of complex engineering systems with long development time (10^+ years), it becomes necessary to embed in the design process the ability to mitigate the risks–cost, schedule, and performance–resulting from requirements changes occurring before fielding the system (see Fig. 3-3). Thus arises the need for process flexibility (see §2.2.2 for a brief discussion of current research on how to provide flexibility during the design process).

3.3 The Need for Flexibility in System Design

Many current engineering systems are required to operate in highly complex and rapidly evolving environments. Significant changes may occur in the market the system is serving, in its economic or strategic environment, as well as in its technology base, thus driving changes in the customers' or users' needs. Market demands and technology levels may vary on time scales significantly shorter than a system's design lifetime.

If the fielded system is not capable of responding to such changes in a timely and cost-effective way, i.e., if the system is not flexible, it is likely to become obsolete (technically and/or commercially) and consigned to storage before the end of its operational lifetime.

As previously stated, the need for flexibility in system design arises for systems that have to operate in highly uncertain environments and/or when the system's design lifetime considerably exceeds the time constants associated with the market dynamics the system is serving in the case of a commercial system and the various technologies embedded in the system. More generally, flexibility should be sought when:

1. The uncertainty in a system's environment is such that there is a need to mitigate market risks, in the case of a commercial venture, and reduce a design's exposure to uncertainty in its environment.

2. The system's technology base evolves on a time scale considerably shorter than the system's design lifetime, thus requiring a solution for mitigating risks associated with technology obsolescence.

60

In other words, flexibility reduces a design's exposure to uncertainty, and provides a solution for mitigating market risks as well as risks associated with technology obsolescence. Figure 3-5 represents a simple schematic of the Innovation and Product Development process that clarifies this statement.
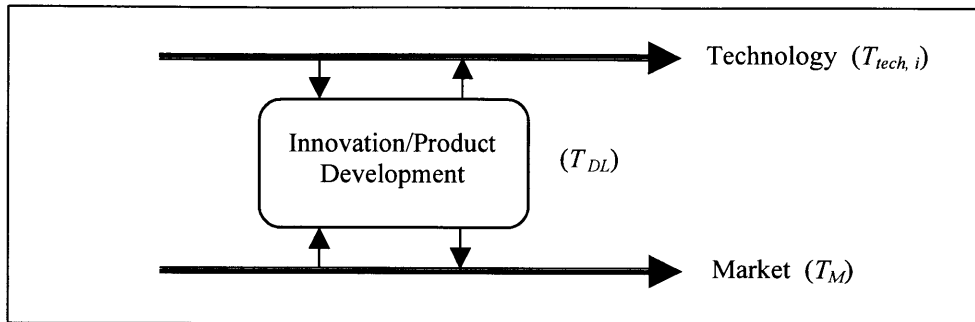
Fig. 3-5. Schematic of the Innovation/Product Development process. Adapted from Allen (1997).

As shown above, the process mediates between two streams of activities:

> One of these [activities] is the development of technological knowledge, or as we more commonly call it, "technology". The other is a developing set of market needs. The basic process is then the matching of information drawn from the two streams: one stream provides market needs; the other provides technological capabilities or potential solutions to meet market needs. Both knowledge of the technology and the market are required [Allen, 97].

For a more elaborate discussion of the Innovation/Product Development process, the reader is referred to Ulrich and Eppinger (1995) on product design and development, and Henderson and Clark (1990) on different types of innovations, in particular architectural innovation versus innovation that affects components and core concepts.

Although simple, Figure 3-5 brings forth the different time scales, or clockspeeds[3], involved in the design process, namely the time constants associated with the rates of change of the product's technology base $(T_{tech, i})$, its environment or market dynamics (T_M), and its design lifetime (T_{DL}).

---

[3] Fine (1998) coined the term "clockspeed" to characterize the rate of change of an industry. "I began to look at [different] industries, seeking to understand their various rates of evolution. I came to think of these rates as industry *clockspeeds*. Each industry evolved at a different rate, depending in some way on its product clockspeed [and] process clockspeed ..."

One particular instance for the need for flexibility in system design, as stated previously, arises when the system's design lifetime considerably exceeds the time constants associated with the market dynamics the system is serving in the case of a commercial system, and the various technologies embedded in the system. Symbolically, this translates into:

$$\begin{cases} T_{DL} >> T_M \\ and \,/\, or \\ T_{DL} >> T_{tech,i} \quad for \; any \; component \; i \end{cases} \tag{3-1}$$

The difficulty of course is in estimating, $a \; priori$, T_M and $T_{tech,i}$. The following paragraphs discuss some of the issues related to market dynamics and technology evolution as well as technology obsolescence, in an attempt to provide an understanding of these elusive–difficult to formally capture and measure–concepts.

3.3.1 Market Dynamics

Market dynamics is a typical example of a (macro) descriptor that makes sense intuitively but is difficult to formally capture and measure. It is evident however that some markets are stable while others undergo rapid changes. The competitive environment, barriers to entry, availability of substitute products, etc. (Porter's Five Forces) drive the dynamics of each market. The following is a typical observation that can be frequently found in business journals or magazines. "Today many new markets are emerging very fast while existing markets are changing rapidly" [Fricke et al., 2000]. The common observation is that different industries can have very different dynamics. Some industries–telecommunications, IT– undergo dramatic changes with astonishing rapidity, while others stroll along at a leisurely pace, undergoing little change, or scarcely being affected by changes occurring elsewhere in the business environment [Fines, 98]. Allen (1997) discusses the "market change" both as a stimuli and a result of technological advances:

Customers' and society's needs change in different ways and at different rates $\left(\dfrac{dM}{dt} \right)$.

Markets vary in their dynamics just as technologies do. Some market niches may be stable, with little change in the requirements from year to year. Other markets are undergoing rapid change. A shift or advance in technology can very often stimulate existing markets or open completely new ones. While there is considerable evidence that

the market provides the stimuli for most commercially successful innovations, technology push has contributed several very important products that have completely changed the markets or created entirely new markets.

The rate of change of a market $\left(\dfrac{dM}{dt}\right)$ introduced by Allen (1997) would provide an indication of the time constant $(T_M)^4$ whose relationship to the system design lifetime (T_{DL}) drives the need for flexibility (see Eq. 3-1). Allen however avoids giving a measure or a procedure for estimating or computing this parameter. Indeed, it is doubtful whether such a parameter can be quantified. The reader interested in a more elaborate discussion of market dynamics is referred to Porter's famous *Competitive Strategy* (1980), or Geroski's *Market dynamics and Entry* (1991). The following example illustrates the need for flexibility (and the consequence of the lack of it) for a system designed to operate in a rapidly evolving market.

3.3.2 Story of Failure: Iridium, Market Dynamics, and the Lack of Flexibility

Iridium is one of the biggest technological gambles and dramatic failures of the commercial space systems. The 66-satellite telephony system entered service in 1998–10 years after it was conceived–and filed for bankruptcy in 1999 after sinking over $3 billions. Lessons from its failure illustrate, among other things, the need for flexibility in high-value assets as discussed in the previous paragraphs. What went wrong?

The target market of Iridium changed between the time the business plan was laid out and when the system became operational. The cellular phone market took off, as did the market for data. By targeting only the market of business travelers, Iridium set itself up against the cellular players. So by the time Iridium entered operation, the cellular-phone technology had overtaken it. The market analysis performed by the system designers identified and explored a steady state or equilibrium configuration of the market; it failed to identify the dynamic nature of its market and did not embed in the system the ability to track a dynamic market and changing customers' needs. Furthermore, its handset was seen to be heavy and outdated, its voice quality not very good, and most importantly, its cost prohibitive compared to other services. In addition, the inability of the Iridium to transfer data proved a serious shortcoming of the system in the age of the Internet. The analysis of the failure of Iridium is a rich topic, yet little explored (lessons from the failure have not yet been extracted, or not yet published). Such an analysis however is beyond the focus of this

---

[4] Given a model of evolution of a market.

work. The purpose of the Iridium example as discussed in this paragraph is to illustrate the relationship between market dynamics and the need for flexibility in system design, as well as the consequence of the lack of flexibility for a system designed to operate in a rapidly evolving market. It also provides a specific example to super-impose on Fig. 3-6 which reads: lack of *flexibility* of a system results in its inability to deal with uncertainty and handle *change*, thus reducing the system's *life span*.
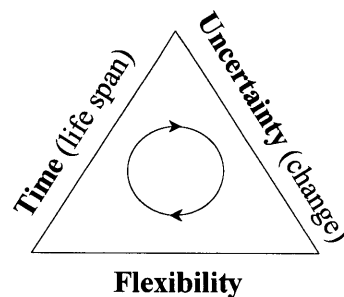
Flexibility

Fig. 3-6. The trilogy: Time, Uncertainty, and Flexibility. Systems that have a longer life span are the ones that are capable of coping with uncertainty and changes in their environment (analyst's perspective). Conversely, if a system is to be designed for an extended design lifetime, the ability to cope with uncertainty and changes has to be embedded in the system (designer's perspective).

3.3.3 Technology Evolution, Obsolescence, and Upgradability: A Problem of Flexibility in System Design

The other stream of activities involved in product design depicted in Figure 3-5 is the development of technological knowledge, or simply the "technology" stream. Many businesses environments no longer have a stable technology base; instead it is novel, or changing rapidly [Iansiti, 98]. It was stated in §3.1 that flexibility should be sought as well when the system's technology base evolves on a time scale considerably shorter than the system's design lifetime, thus requiring a solution for mitigating risks associated with technology obsolescence. An understanding of technology evolution, obsolescence, and the technology upgrade problem is therefore required in order to recognize flexibility as a solution for mitigating risks associated with technology obsolescence. These issues are discussed in the following paragraphs.

Many complex systems, particularly aircraft and spacecraft as well as major military systems, often have design lifetime far in excess of several life spans of many of their supporting technologies and components. Indeed the commercial new product cycle for many technologies, including the rapidly evolving electronic component industry, is two years or less. With systems

that are designed to remain in operation for 15 years or more, obsolescence problems become critical [Hitt and Schmidt, 98].

Although increased reliability as well as budgetary constraints have lengthened complex systems life cycles, rapid advances in technology have dramatically shortened component life cycle. This decrease is particularly acute for electronic components but affects non-electronic components as well. The disparity between the long life cycles of systems and the short life span of components embedded in the system requires a careful consideration of issues of obsolescence and upgradability, or the ability to modify a system after it has been fielded in a timely and cost-effective way. This is precisely the characteristic feature of flexibility of a design.

Component obsolescence affects all programs, both commercial and military, and is a major consideration in today's weapon systems. Examples of technology obsolescence and upgrade for instance include the replacement of rotating gyros with ring-laser or fiber-optic gyros, or the use of flat panel displays (FPD) to replace cathode ray tubes (CRT). Flat panel displays (FPDs) are thin electronic devices used to create visual images or present textual information. The dominant FPD technology today is the liquid crystal display or LCD. LCD technology is supported by a growing manufacturing base, and sales yield a multi-billion dollar revenue stream. FPD/LCD offer substantial advantages in terms of size, weight, durability, and performance over the aging cathode-ray tubes or CRT, which have become increasingly difficult and expensive to acquire [Lippitz, 99].

The best-known aspect of technology obsolescence is non-availability of parts as vendors move on to newer technologies and products [Borky et al., 98]. For example, Boeing's 777 relies on Intel 80486 microprocessors for flight management. However this processor is no longer being manufactured by Intel, which has redirected its resources to the production of the Pentium family. Furthermore, Intel is not interested in developing custom products for avionics for instance because at less than 1%, the market is insignificant. This leaves Boeing with an difficult dilemma: Should it try to continue to procure the outdated parts by using either warehoused or new versions of the discontinued part at up to 20 times the original price, or should it attempt a re-design (at what cost?) to take advantage of newer technology? [McKluskey and Das, 98]

As stated above, obsolescence problems affect all programs, both commercial and military. Luke *et* al. (1998) for instance affirm that every aircraft in the U.S. military inventory today has problems with non-availability of electronic components. The following paragraph attempts to quantify a design's exposure to the risks of components' obsolescence.

3.3.3.1 Time to Obsolescence and the "Hastings Vector" of a Design

Figure 3-7 shows a typical component's progress towards obsolescence.

Emerging	Growth	Maturity	Saturation	Decline	Phase Out	Discontinued

	Emerging	Growth	Maturity	Saturation	Decline	Phase Out
Sales	Slow Increase	Rapid Increase	Stable	Leveling Out	Decreasing	Scarce
Price	High	Declining	Stable	Stable	Rising	High
Usage	Low	Increasing	Stable	Stable	Decreasing	Phasing Out
Profits	Low	Increasing	Stable	Stable	Reasonable for Survivors	X
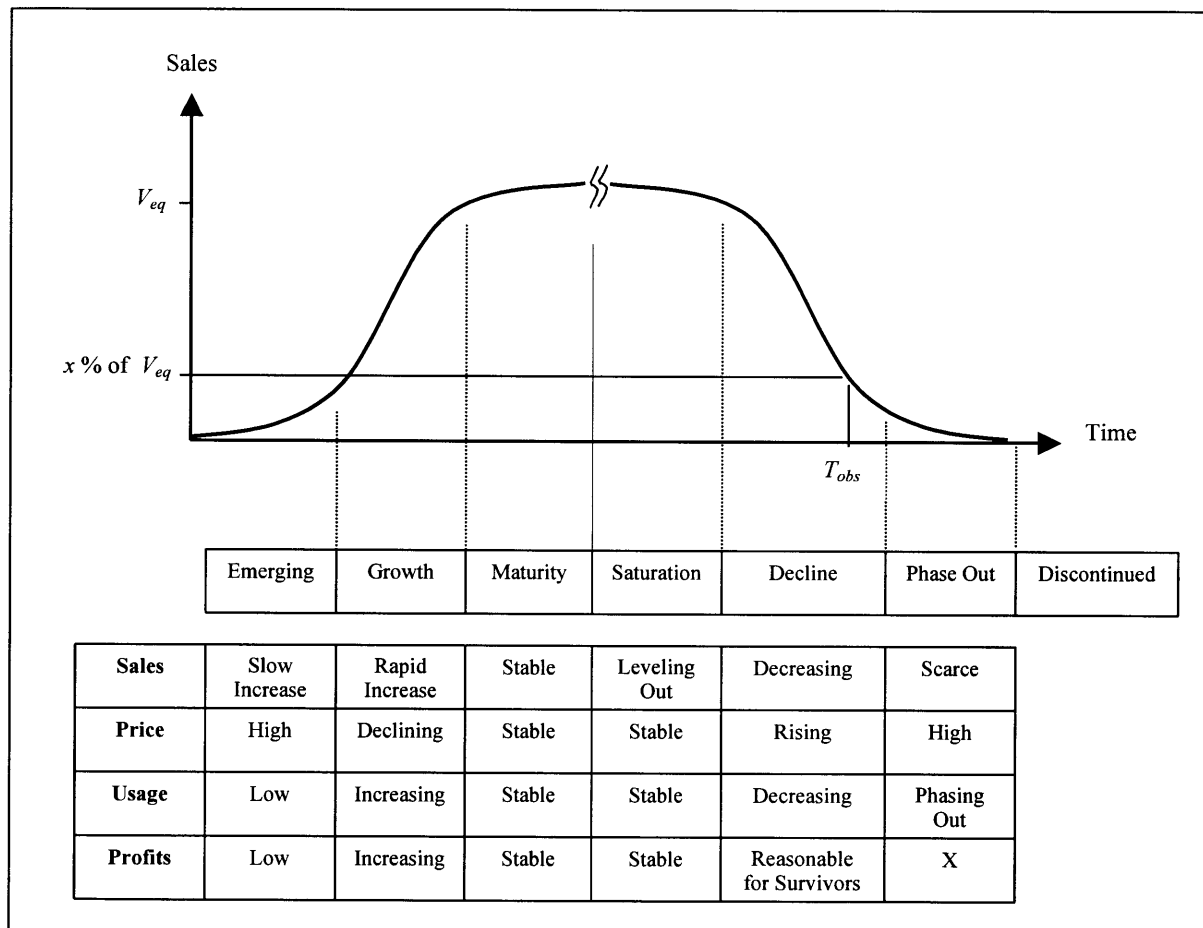
Fig. 3-7. Typical component life cycle and phases. Adapted from the ANSI/EIA-724-97 Product Life Cycle Data Model [Hatch, 00]. Time to obsolescence T_{obs} can be defined in relation to sales volume.

One way of defining Time to Obsolescence T_{obs} is in relation to sales volumes, as illustrated above. T_{obs} of a component can be treated as a random variable with an associated distribution function $P(T_{obs})$ and a probability density function $p(T_{obs})$ such that:

66

$$P(0 < T_{obs} < \tau_2) = \int_0^{\tau_2} p(t)dt \qquad\qquad p(t) = \frac{d}{dx}P(t) \qquad\qquad (3\text{-}2)$$

What distribution function should be used for T_{obs}? There are a handful of parametric models that can be used as probability density functions for T_{obs}. Evans *et* al. (2000) discuss 40 different distribution functions. A distribution function is preferred over another for a given set of data when (i) there is a physical/statistical argument that theoretically explains the data (ii) a particular model has been used successfully for a similar phenomenon (iii) the model provides a good empirical fit to the data. However, given the estimation nature of the problem discussed here, i.e., T_{obs} of a component is being estimated when the component is selected for a design, no data is available and the choice of a distribution function for T_{obs} is left to the discretion of the analyst/designer. In the following, the lognormal distribution will be used. The lognormal distribution, like the Weibull distribution, is a very flexible model that can empirically fit many types of data and phenomena. It has two parameters–unlike the Rayleigh distribution for instance which has only one parameter fixing simultaneously the mean and the variance–the median m, and the standard deviation σ of *log(X)*:

$$p(X) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)X} Exp\left[-\left(\frac{\log(X/m)}{\sigma\sqrt{2}} \right)^2 \right] \qquad\qquad (3\text{-}3)$$

The lognormal distribution is applicable to random variables that are constrained by a lower bound, and can have few large values. A third parameter τ, often called waiting time or shift parameter, defines the lower bound of the random variable. Since each component used in a system design follows a life cycle of its own and progresses toward obsolescence at its own rate, an index i is used to differentiate between various components. For a component i, the probability density function of its Time to Obsolescence is modeled as follows:

$$p_i(\widehat{T}_{obs}) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)\left(\widehat{T}_{obs} - \tau\right)} Exp\left[-\left(\frac{\log\left(\left(\widehat{T}_{obs} - \tau\right)/m\right)}{\sigma\sqrt{2}} \right)^2 \right] \qquad\qquad (3\text{-}4)$$

In some cases, it is easier to obtain–or understand–the cumulative distribution function of a random variable than its probability density function. Figure 3-8 presents both a cumulative

distribution function and probability distribution function of the Time to Obsolescence for a typical microprocessor (m=1.5 years, σ = 0.8 years, τ = 0.5 years).
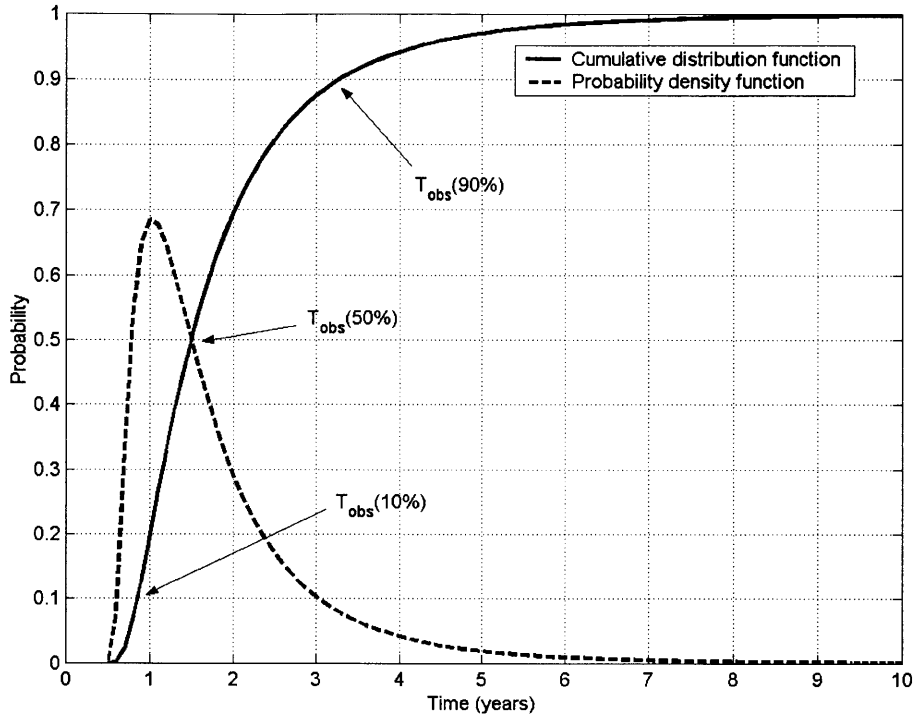
Fig. 3-8. Cumulative distribution function and probability density function of the Time to obsolescence for a typical microprocessor (m=1.5 years, σ = 0.8 years, τ = 0.5 years).

Let $C = \{c_i\}$ be the set of components having direct impact on mission performance, and $p_{i,t}(\widehat{T}_{obs})$ the probability density function of the Time to Obsolescence, available at time t, of component i from the set C:

$$C = \{c_i\} \qquad \rightarrow \qquad \left\{ p_{i,t}\left(\widehat{T}_{obs}\right)\right\} \qquad\qquad (3\text{-}5)$$

For $t = T_{ops}$ (time of fielding the system), we can define the Hastings vector of a system with a design lifetime of T_{DL} as follows:

$$H_{x\%} = \left[\frac{\widehat{T}_{obs_1}(x\%)}{T_{DL}}, \frac{\widehat{T}_{obs_2}(x\%)}{T_{DL}}, \dots, \frac{\widehat{T}_{obs_n}(x\%)}{T_{DL}} \right]^T \qquad\qquad (3\text{-}6)$$

H characterizes the exposure of the design to problems with component obsolescence, starting from the time the system is fielded. A system is unaffected by obsolescence problems if $H_i \approx \theta(1)$, i.e., is of the order of 1, for all *i*. In other words, all components have similar Time to Obsolescence as the system's design lifetime (see Eq. 3.6). When this is not the case, *H* allows program managers to identify problem components that are likely to become obsolete early in the system's operational life, i.e., for which $H_i \ll 1$, and to take actions in order to mitigate the risk of early component obsolescence.

For instance in the case of the Boeing 777 and the Intel 80486 in its Flight Management System, it is well known that Intel introduces major product improvements on the market every 16 to 24 months. The price of the current product is reduced when the new product is introduced; then as the market assimilates the new product, the older product is phased out typically in 3 to 4 years. Assuming the 777 will remain in service for 30 years, then:

$$H_{\mu p} \approx \frac{3}{30} = 0.1$$

It is therefore clear that the processor will become obsolete during the aircraft's operational lifetime (assuming all else remaining constant, the processor will be dragging ten generations behind from the technology leading edge by the aircraft is consigned to storage). Upgrade opportunities therefore will become available that offer improved or new functionality. The aircraft's Flight Management System (FMS), therefore, should be designed in a way to accommodate changes in a timely and cost-effective way, i.e., flexibility should be embedded in the design of the FMS.

3.3.3.2 Obsolescence Prediction and Management

The above example points out to the necessity of performing technology obsolescence predictions during a system's development phases–when parts and components are selected–in order to evaluate *H* and assess the criticality of obsolescence related problems. Several commercial companies offer obsolescence management systems and provide analysis for component obsolescence vulnerability and component life cycle projection[5]. Projections are made for which components in a design will become obsolete in the near future, what is the current component

---

[5] See for example TACTech at http://www.tactech.com, or Manufacturing Technology Incorporated (MTI) at http://www.mtifwb.com.

availability, where procurement problems exist, what are the replacement options, etc. For example, TACTech reports that the semi-conductor technology baseline changes on average between 9 months and 8 years (see Table 3-1).

Table 3-1. Semiconductor technology base Time to Obsolescence.

Average Time to Obsolescence for New Generations of Commercial Integrated Circuit	
Logic families	6 years
Memory families	9 months
Microprocessors	2 years
DSP	3 years
Linear/Interfaces	8 years
Gate arrays	2 years

Obsolescence problems in Commercial-Off-The-Shelf (COTS) equipment are inevitable. While all COTS equipment are subject to obsolescence, particular component classes or parts are prone to specific problems, ranging from minor to volatile. The term volatile refers to frequent and likely changes. The volatile category includes software, central processing units (CPUs), memory chips, etc. Graphic displays and keyboards are less volatile [ARINC, 00]. The H vector introduced above captures this classification and quantifies a component's exposure to obsolescence during the system's operational lifetime in which it is integrated. Figure 3-9 illustrates different levels of system integration.

```
┌─────────────────────────────────────────────────────┐
│ Mission  ·······································  ┊·········· User
│   ┌───────────────────────────────────────────┐ ┊
│   │ System/Platform ························  ┊···· System manager/integrator
│   │   ┌───────────────────────────────────┐ ┊ ┊
│   │   │ Subsystem  ·····················  ┊·┊···· Subsystem integrator
│   │   │   ┌───────────────────────────┐ ┊ ┊ ┊
│   │   │   │ Component ···············  ┊·┊·┊···· Supplier
│   │   │   └───────────────────────────┘ ┊ ┊
│   │   └───────────────────────────────────┘
│   └───────────────────────────────────────────┘
└─────────────────────────────────────────────────────┘
```
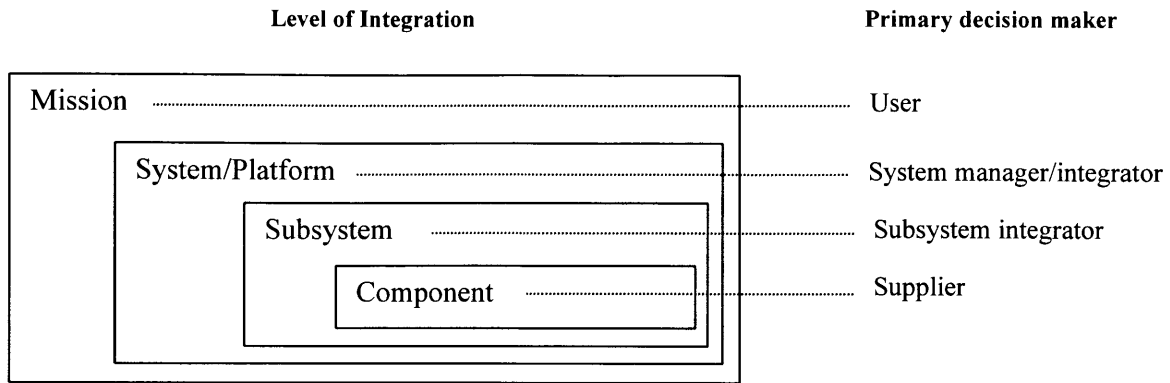
Fig. 3-9. Levels of Integration and upgrade management decision makers. Suppliers produce components (e.g., FPD). Subsystem integrators incorporate these components into subsystems (e.g., display units). Contractors or system managers/integrators oversee the integration of subsystems into platforms or systems, such as an aircraft, a ship, a ground vehicle, etc. These systems are then delivered to users who control how systems are used to perform missions [Lippitz, 99].

Once H is evaluated, a plan should be devised to mitigate the risks associated with particular components obsolescence. The Defense Micro-Electronics Activity (DMEA[6]) for instance defines three levels of practices to mitigate the impacts of technology obsolescence and diminishing manufacturing sources and material shortages (DMSMS). There are:

Level 1–Practices are implemented to resolve current obsolete items. Some of these activities may be considered reactive (e.g., parts list monitoring, supportability checklist).

Level 2–Minimal required practices are needed to mitigate the risk of future obsolete items. The majority of these activities are considered proactive (e.g., obsolescence prediction, awareness training, DMSMS solution database).

Level 3–Advanced practices are required to mitigate the risk of obsolescence when there is a high opportunity to enhance supportability or reduce total cost of ownership. These proactive activities may require additional funding up-front in the design process [ARINC, 00].

---

[6] The primary mission of the Defense Micro-Electronics Activity (DMEA), Department of Defense executive agent for micro-electronics Diminishing Manufacturing Sources and Materials Shortages (DMSMS), is to "leverage the capabilities and advantages of advanced technology, reduce Operating and Support (O&S) costs, and reduce the effects of DMSMS." [ARINC, 00]

Practices that fall under the DMEA Level 3 enable an electronic component or subsystem to be designed in a way easily changed. These are a subset of the flexibility enabling practices discussed in the following chapter.

3.4 The DoD's Exposure to Technology Obsolescence Problems

The DoD defines obsolescence as diminishing manufacturing sources and material shortages. DMSMS is a serious issue for the DoD, the airline community, and many commercial industries. Diminishing Manufacturing Sources and Material Shortages (DMSMS) concerns the loss or impending loss of manufacturer or suppliers of critical items and raw material due to discontinuance of production. DMSMS can be caused by rapid changes in item or material technology, uneconomical production requirements, federal environmental or safety requirements, and limited availability or increasing cost of item and raw materials [ARINC, 00].

In the DoD, concern is growing about the cost of resolving current and future technology obsolescence/DMSMS problems. The Deputy Under Secretary of Defense for Logistics (DUSD (L)) indicates that the average cost to redesign a circuit card to eliminate obsolete components is $250,000. Similarly, the Electronic Industry Association (EIA) Manufacturing Operations and Technology Committee report a cost range to redesign of between $26,000 and $2 million [ARINC, 99].

Furthermore, with decreasing defense dollars available to purchase new weapon systems, the inventory of existing systems will have to last many more years than originally planned. As the avionics for instance of these aging systems get older, they become more expensive to maintain due to component obsolescence. In addition, expanding missions and changing requirements (flexibility) lead to growth in the embedded software, which in turn, requires additional processing and memory capacity. Both factors, **parts obsolescence** and new **processing capacity**, result in the need to replace the old computer hardware with newer, more capable microprocessor technology [Luke et al., 98].

Figure 3-10 is a graphic approximation of the current DoD budget breakdown allocated to each phase of the procurement life cycle. It clearly shows that the biggest portion goes to the Operations and Support (O&S) segment [Fitzhugh, 98].
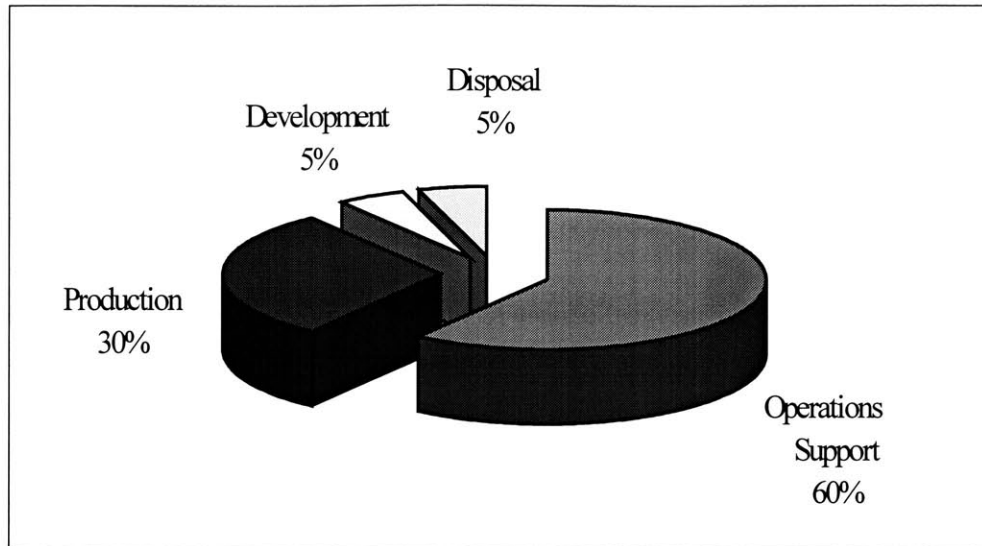
Fig. 3-10. Approximation of the current DoD percentages budget allocated to each element of the procurement life cycle. Adapted from [Fitzhugh, 98].

Although they have always been important, Life Cycle Costs (LCC) have become particularly important at the DoD. While budgets have declined over the last decade, Operations and Support (O&S) costs have remained almost constant, while the bulk of the budget reductions were absorbed in the procurement area. Given that the procurement budget has long been closely managed, the O&S segment, which currently dominates the military's budget, represent the greatest potential for cost savings [Fitzhugh, 98].

While there are several costs drivers associated with O&S, one of the major contributors is parts components obsolescence.

"[Obsolescence] occurs when the last known manufacturer or supplier of an item or of a raw material gives notice that they intent to cease production. The majority of these cases have been in the electronics area, however obsolescence problems affect all weapon systems and material categories. Obsolescence may occur at any phase in the acquisition cycle [of a weapon system], from design and development, through post-production, and have the potential to severely impact weapon systems supportability and Life Cycle Costs." [Fitzhugh, 98]

Because of this budgetary situation, planners have been forced to re-evaluate the traditional process of acquiring and maintaining complex weapon systems. It is however the up-front in the design process that the opportunity for mitigating future obsolescence risks promises the best

73

return on investment, where particular emphasis should be placed on the system's flexibility or its ability to be modified in a timely and cost-effective way in order to accommodate a different/new set of requirements.
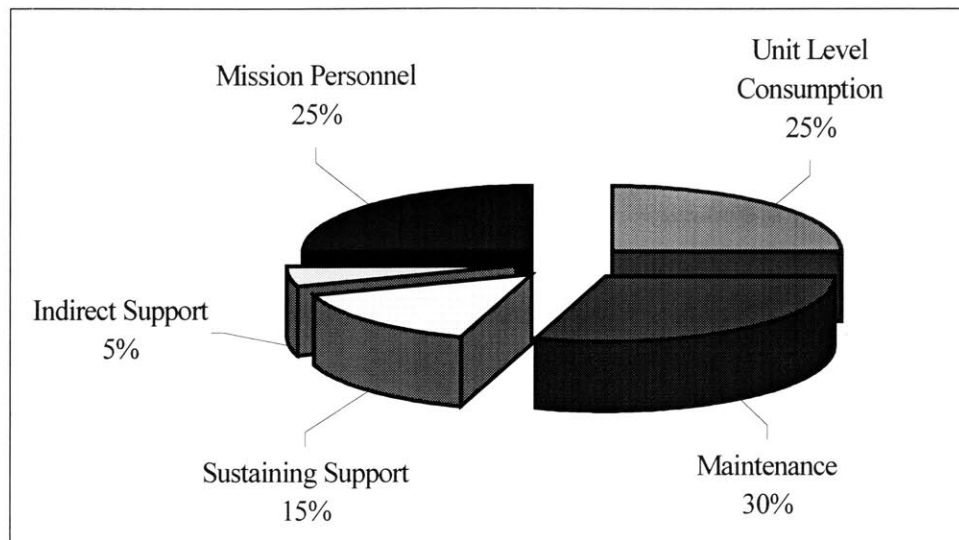
Fig. 3-11.Operations and Support (O&S) percentage cost structure breakdown. Adapted from [Fitzhugh, 98].

3.5 Summary

The previous chapter reviewed the concept of flexibility as discussed in various fields of investigations and extracted its characteristic features. Flexibility in this work in defined as the property of a system that allows it to respond to changes in its initial objectives and requirements–both in terms of capabilities and attributes–occurring after the system has fielded, in a timely and cost-effective way.

The chapter identified and characterized the situations in which flexibility in system design is needed. In particular, it was argued that flexibility should be sought when:

1. The uncertainty in a system's environment is such that there is a need to mitigate market risks, in the case of a commercial venture, and reduce a design's exposure to uncertainty in its environment.

2. The system's technology base evolves on a time scale considerably shorter than the system's design lifetime, thus requiring a solution for mitigating risks associated with technology obsolescence.

Section 1 motivated the subject from a specific perspective: the need for flexibility in high-value on-orbit assets. Section 2 recalled the distinction between process flexibility and flexibility of a design, and discussed the dynamics of system requirements during the requirement generation phase. This in turn is used to make a brief statement on the need for process flexibility. Several examples are provided in support of this statement. Section 3 is the center of this chapter; it addresses the need for flexibility in system design. First the case of flexibility is made in relation to fast market dynamics (on a time scale of the order of the system design lifetime). The failure of the Iridium system is discussed as an example of the consequences of the lack of flexibility of a system designed to operate in a rapidly evolving market. Second, it is argued that flexibility should be sought when the system's technology base evolves on a time scale considerably shorter than the system's design lifetime. A discussion of technology evolution and obsolescence is provided, and a metric is introduced to quantify the disparity between components' life cycle (or Time to Obsolescence) and the system's design lifetime, thus driving the need for flexibility. The DoD's exposure to technology obsolescence problems is touched upon in Section 4.

As discussed earlier, a comprehensive treatment of flexibility in system design should address the following questions:

1. What is flexibility?
2. Why or when is flexibility needed in system design?
3. How can one design for flexibility? What are the design principles for embedding flexibility in system design?
4. What are the trade-offs associated with designing for flexibility? What is the value of flexibility and what are the penalties (cost, performance, risk, etc.), if any, associated with it?

This chapter focused on the second question and identified the situations in which flexibility in system design is needed. The following chapters expand on the subsequent questions in the particular case of spacecraft design.

Part II

Spacecraft Design Lifetime and On-Orbit Servicing

The first part of this thesis discussed issues of flexibility in system design in the general case. Namely, the two following questions were addressed. 1) What are the characteristic features of flexibility in system design? 2)What drives the need for flexibility in system design, and what are the attributes of an environment in which flexible designs should be sought and fielded?

One way through which flexibility manifests its criticality to systems architects is in the specification of the system design lifetime requirement. Part II addresses issues of design lifetime, and ways to provide and value flexibility, in the particular case of space systems.

Chapter 4

On Spacecraft Design Lifetime

"In God we trust. All else show us the data."

Anonymous.

4.1 Introduction

In recent years, several space programs have chosen to increase their space segment design lifetime. Over the last two decades, telecommunications satellites for instance have seen their design lifetime increase on average from seven to fifteen years. This trend is also observed in the design and development of many high value assets, e.g., the average lifetime of a helicopter delivered today can exceed thirty years or 20,000 hours of operation.

In most cases, increasing the space segment design lifetime was driven by the desire to maximize the return on investment. However, extending the design lifetime has several side effects. First, doing so leads to larger and heavier satellites as a result of several factors such as additional propellant for orbit and station-keeping, power generation and storage. This additional mass in turn increases the satellite's development and production costs. Second, as the satellite design lifetime increases, the likelihood that the satellite becomes obsolete, technically and commercially, before the end of its mission increases. In many cases, the initial circumstances from which the original system requirements were derived change or are modified during the spacecraft's operational lifetime. Setting a spacecraft design lifetime requirement therefore can be a critical task for system designers. But what drives a spacecraft design lifetime? How do designers, managers, and or customers decide on spacecraft design lifetime? Are there any economical considerations such as minimizing cost-per-operational day or maximizing the return on investment (ROI) for establishing this requirement, or is the design lifetime mainly dictated by technical limitations? Some of the technical considerations that limit a spacecraft lifetime include

the depletion of consumables, the degradation due to spacecraft/environment interaction (of solar panels, radiators, electronics, etc.), wear and tear (thrusters pulse life, wheel bearings, etc.), and reliability/redundancy issues.

Consider for instance AT&T's Telstar 3 communications satellites based on the Hughes HS-376 bus. These satellites have ten-year design lives, a significant increase over the seven-year lives for earlier satellite models. Life extension was made possible by the use of improved Nickel-Cadmium batteries and the introduction of solid-state power amplifiers in place of traveling wave tubes.

Questions regarding the design lifetime requirement can fall into the three following categories: 1) What limits the design lifetime? 2) How does the total system mass and cost change as a function of the design lifetime requirement? 3) What does the customer ask the contractor to provide for a spacecraft design lifetime, and why?

Although related, these three questions nevertheless cover different realities. The first question addresses the issue of the lifetime "boundary": How far can designers push a spacecraft's design lifetime and why can't they extend it any further? The technical considerations listed above dictate this boundary.

The second question, closely related to the first one, focuses on the effects of varying the design lifetime requirement on the total spacecraft mass and cost, or the cost to Initial Operating Capability (IOC). It is clear that the design lifetime will have a strong bearing on the mass and cost of the system by affecting the power budget (Beginning Of Life—BOL—solar panel, battery capacity, etc.), the propellant budget (orbit and station keeping), the level of redundancy, and other key system parameters.

The third question builds on the two previous ones: Given the maximum achievable design lifetime, as well as the impact of the mission duration on the spacecraft mass and cost, what does the customer require for spacecraft design lifetime? The design lifetime should not necessarily be set to the maximum achievable value. For example, a commercial customer may not want to make the contract life of a spacecraft too long. New or enhanced payload capabilities, e.g., better spatial resolution for an optical instrument, might be developed and become available within a couple of years following deployment, hence the need to launch a new satellite or risk losing market share to a competitor who launches later with newer and more advanced capabilities.

The answer to the first question can be easily obtained by examining the maximum operational lifetimes of the various technologies embedded in the spacecraft. In this Chapter, we investigate Question 2, that is: How does the spacecraft lifetime requirement impact the design and sizing of the various subsystems? How do these subsystems scale with the design lifetime, and consequently, how does the total spacecraft mass and cost scale with the design lifetime requirement? Answering this question is pivotal for understanding the rationale in specifying the lifetime requirement; hence it is a prerequisite for answering Question 3.

This Chapter is organized as follows: First, typical percentage mass contributions of different subsystems to the total spacecraft mass are presented, and the main mass drivers are identified. Second, the impact of the design lifetime requirement on the sizing of the spacecraft subsystems is investigated. The results are then integrated and typical spacecraft mass profile as a function of the design lifetime are presented. These mass profiles are in turn transformed into cost profiles, and the *cost-per-operational day* metric is introduced. Finally, the limitations of the previous analysis are addressed and the implications of this analysis are discussed.

4.2 Typical Mass Distribution of Satellites

In order to assess the impact of the design lifetime on the spacecraft mass, it is useful to identify mass contributions of the subsystems to the spacecraft total mass. Table 4-1 shows some historical spacecraft mass distribution data. For example, the electrical power subsystem (EPS) accounts on average for 30% of satellite dry mass, with a standard deviation of 7%. The EPS, along with the payload and spacecraft structure, are the major mass contributors and make up approximately 80% of satellite dry mass.

Table 4-1: Percentage mass distribution, averages and standard deviations. Adapted from [Wertz and Larson, 99].

	EPS	Payload	Structure	ADCS	TT&C	Propulsion	Thermal
			Percentage of satellite dry mass (standard deviation)				
Communication	32 % (5)	27% (4)	21 % (3)	7 % (2)	5 % (2)	4 % (1)	4 % (2)
Navigation	32 % (3)	21 % (2)	23 % (3)	6 % (0.5)	5 % (1)	3 % (0.5)	10 % (1)
Remote sensing	25 % (4)	36 % (5)	20 % (3)	5 % (2)	4 % (1)	7 % (3)	3 % (1)
Average	30 % (7)	28 % (7)	21 % (5)	6 % (3)	5 % (2)	4 % (3)	6 % (2)

4.3 Spacecraft Subsystems and Design Lifetime

We now examine how different subsystems scale with the design lifetime requirement. This requirement is a key parameter in sizing several spacecraft subsystems: it directly impacts the design and sizing of some subsystems, e.g., the electrical power subsystem, and indirectly impinges on others, e.g., the structure. These influences and coupling are qualitatively captured in Table 4-2. The diagonal in Table 4-2 represents the direct impact of the design lifetime requirement on each subsystem. The off-diagonal terms read as follows: Subsystems in the first column scale with the design lifetime, driven by changes in subsystems in the first row. The number of crosses represents the degree of influence (+++ major influence, + minor influence).

Table 4-2. Design lifetime influence matrix.

	ADCS	TT&C	EPS	Thermal	Structure	Propulsion	Propellant
ADCS			+ + +	+	+ + +	+	+ +
TT&C		Redundancy, shielding					
EPS			Solar array degradation, batteries' DOD				
Thermal			+ + + ($\Delta P_{BOL\text{-}EOL}$)	Degradation of thermal properties of coating			
Structure	+	+	+ + +	+		+	+ +
Propulsion						Wear-and-tear/On-Off cycles	
Propellant	+	+	+ + +	+	+ + +	+	Increase in ΔV with design lifetime

4.3.1 Electrical Power Subsystem

The electrical power subsystem generates power, conditions and regulates it, stores it for peak demand or eclipse operation, and distributes it throughout the spacecraft [Wertz and Larson, 99]. The design lifetime is a key parameter in sizing the EPS. It directly impacts 1) the life degradation of the solar arrays, hence their surface, and consequently their mass, and 2) the battery capacity through the extended number of cycles, hence reduced Depth-Of-Discharge (DOD) with design life. The design lifetime also indirectly impacts the sizing of the power controllers and regulators as well as the harnesses and cabling that interconnect the spacecraft subsystems. In this subsection, we explore how the EPS scales with the design lifetime.

Solar arrays: "In designing solar arrays, experts typically trade mass, surface, and cost [...] Life degradation L_d of solar arrays occurs because of thermal cycling in and out of eclipses, micrometeoroid strikes, plume impingement from thrusters, material outgassing, and radiation damage throughout the duration of the mission." [Wertz and Larson, 99]. Life degradation is a function of the design lifetime and can be estimated as follows:

$$L_d = (1 - degradation/year)^{T_{Life}}$$

(4-1)

The *degradation-per-year* is a function of the spacecraft orbital parameters (location with respect to the Van Allen belts) as well as the solar cycle. Typically, for a Silicon solar array in LEO, power production can decrease by as much as 3.75% per year, and 2.75% for Gallium-Arsenide [Wertz and Larson, 99].
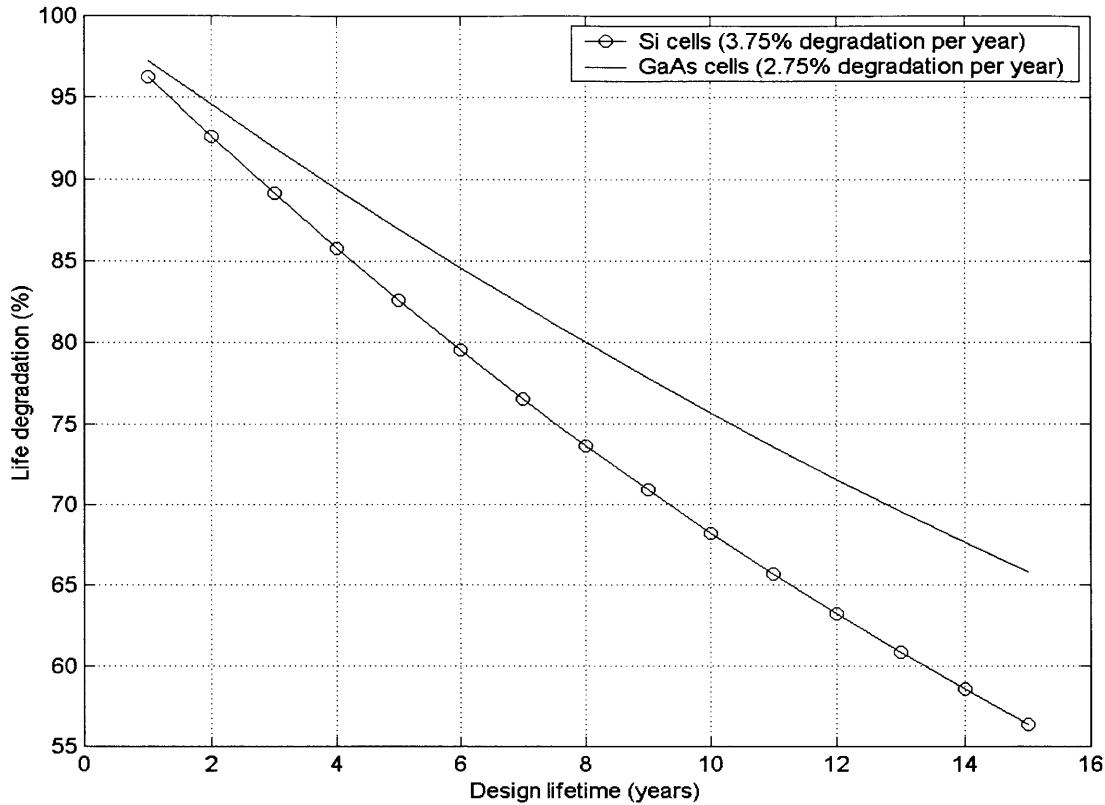
Fig. 4-1. Typical life degradation of solar arrays in LEO as a function of design lifetime.

Figure 4-1 shows typical life degradations of Silicon solar arrays and Gallium-Arsenide arrays as a function of design lifetime. Gallium-Arsenide cells are both more efficient (19% energy conversion efficiency) and degrade slower than Silicon cells (efficiency of about 15%). For instance, given a six years design lifetime, the power output of Silicon arrays will degrade by 80%, and 85% for Gallium-Arsenide arrays. The array's performance at the end-of-life is given by (P_{BOL} = Power at beginning-of-life, W, and P_{EOL} = Power at end-of-life, W):

$$P_{EOL} = P_{BOL} \times L_d \qquad (4\text{-}2)$$

Given a power requirement at end-of-life, the power output of the solar arrays at beginning-of-life scales inversely with life degradation L_d and the solar arrays have to be over-designed to accommodate this performance degradation. Figure 4-2 illustrates the relationships between the P_{BOL} and the design lifetime for a 4KW P_{EOL} requirement. It reads as follows: in order to deliver 4KW for instance at the end-of-life of a ten-years mission, the solar arrays should be designed to provide approximately 5.5KW in the case of GaAs cells and 6KW in the case of Si cells at the beginning-of-life.

82
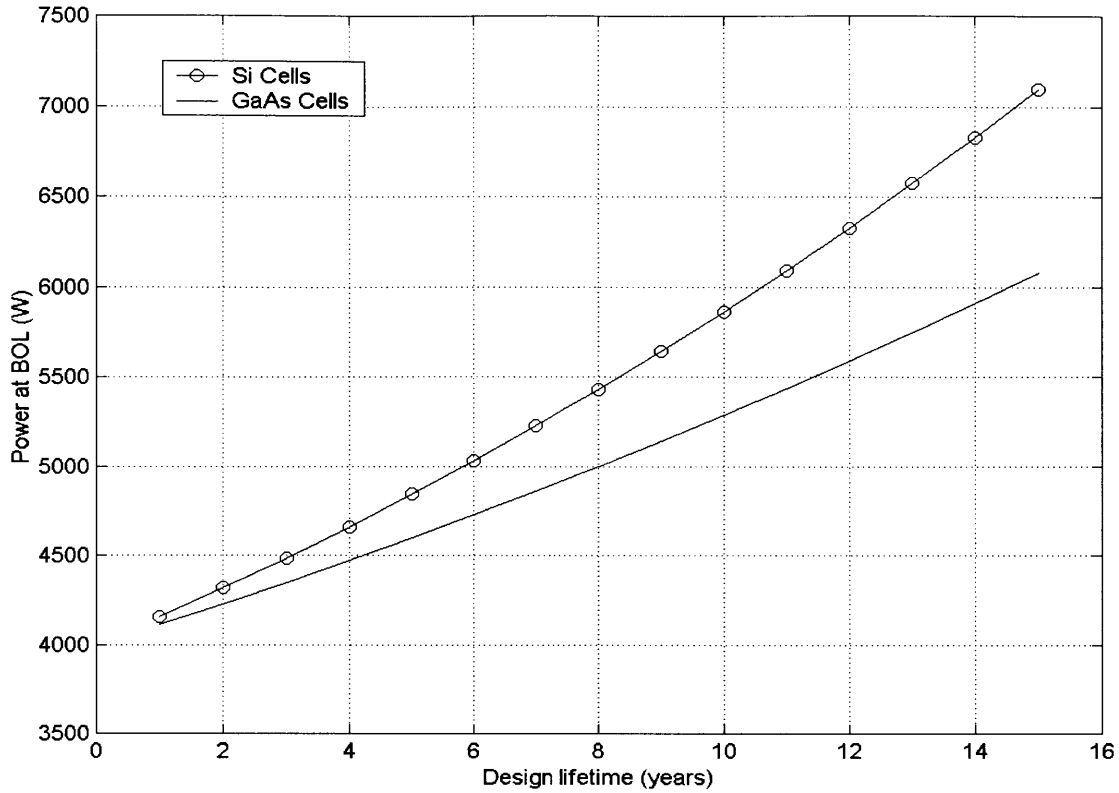
Fig. 4-2. Solar array P_{BOL} as a function of design lifetime for a 4KW P_{EOL} requirement.

The solar array surface required to produce the P_{BOL} is approximately:

$$S_{sa} \cong \frac{P_{BOL}}{I_s \times \eta} \qquad (4\text{-}3)$$

I_s = Solar intensity at 1AU, 1367W/m^2, and η = Solar cells energy conversion efficiency (19% for GaAs and 15% for Si cells) times array inherent degradation.

Given the specific performance of the array in W/kg (or W/m^2), the mass of a planar array is directly evaluated. Typical specific performances range between 20W/kg and 70W/kg. Results for a nominal specific performance of 40W/kg are presented in Figure 4-3.
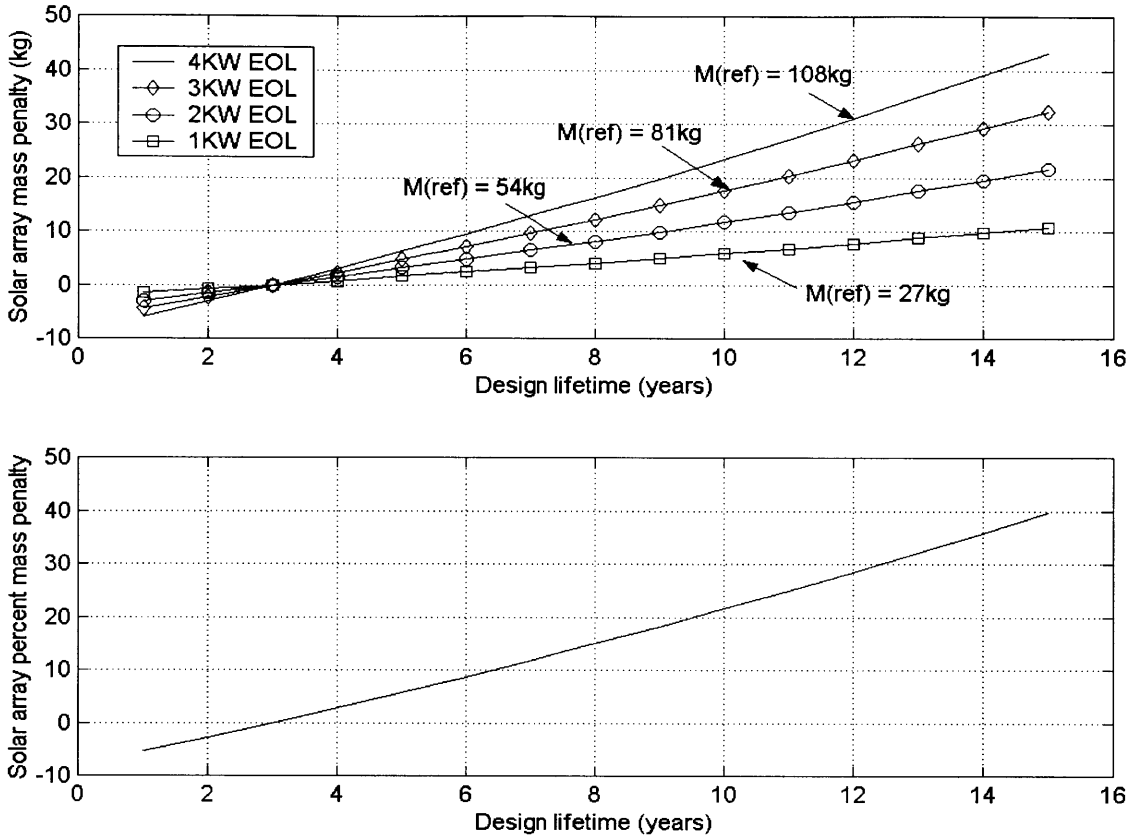
Fig. 4-3. Solar array (GaAs) mass, mass penalty, and percent mass penalty as a function of the design lifetime. The reference mission is 3 years.

In order to deliver 4KW for instance at the end-of-life of a 10-year mission, the solar arrays would weigh approximately 130kg, that is 22kg in excess of a solar array delivering the same 4KW at the end-of-life of a 3-year mission. This is equivalent to approximately 20% mass penalty for 7 extra years of life.

Batteries: Spacecraft in Earth orbit undergo between 90 eclipses and 5500 eclipses per year. The former figure is typical of a GEO satellite, the latter a satellite in LEO. During eclipse, electric power is supplied by secondary batteries that are recharged by the solar arrays when the spacecraft re-emerges into sunlight. In addition, there are some instances when batteries are called upon to provide peak power in sunlight periods. The existing state-of-the-art and space-qualified batteries (Nickel-Hydrogen) are heavy and can constitute up to 15% of the dry mass of a typical communications satellite[2]. Current secondary battery technology includes Nickel-Cadmium, which is a very common space qualified secondary energy storage system. Nickel-Hydrogen batteries are currently the energy storage system of choice for most aerospace

applications where high specific energies and long design life are required. Lithium-Ion and Lithium-Carbon batteries currently under development with expected space qualification for GEO and LEO applications by 2005-2010 [Wertz and Larson, 99]. Lithium-Ion and Lithium-Carbon batteries would offer significant mass and volume reduction compared to Nickel-Cadmium and Nickel-Hydrogen technology, as illustrated in Figure 4-4.
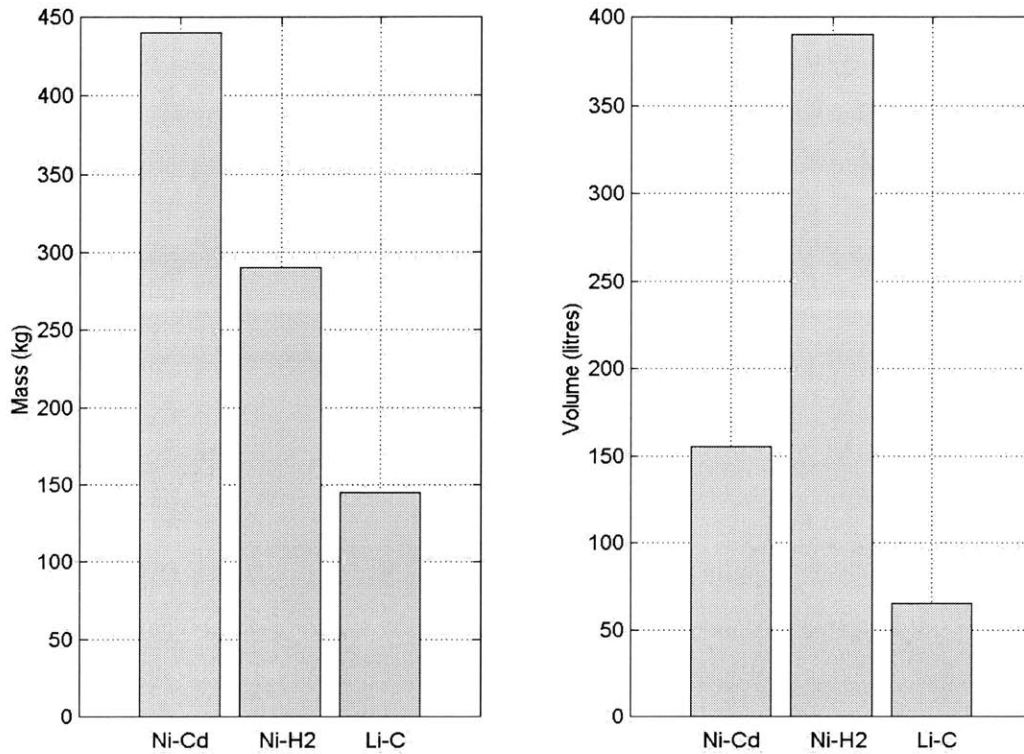
Fig. 4-4. Mass and volume of different types of batteries for a 10 KWh capacity [Dudley and Verniole, 97].

The design lifetime significantly impacts the sizing of secondary batteries. Indeed the amount of energy available from secondary batteries, the depth-of-discharge or DOD, decreases with the number of cycles of charging and discharging. To first order this number of cycles is equal to the number of eclipses a satellite encounters during its design lifetime. Typically a satellite in GEO undergoes two periods of 45 days with eclipses lasting no more than 72 minutes, hence 90 cycles of charging and discharging per year. Satellites in LEO undergo approximately one eclipse per orbit. For a 90-minute orbit, this amounts to 16 eclipses per day, or approximately 5500 cycles per year, with a maximum shadowing period of nearly 36 minutes per orbit. Figure 4-5 represents the DOD as a function of the number of cycles a battery undergoes charging and discharging, as well as the DOD as a function of the design lifetime of a satellite in GEO.
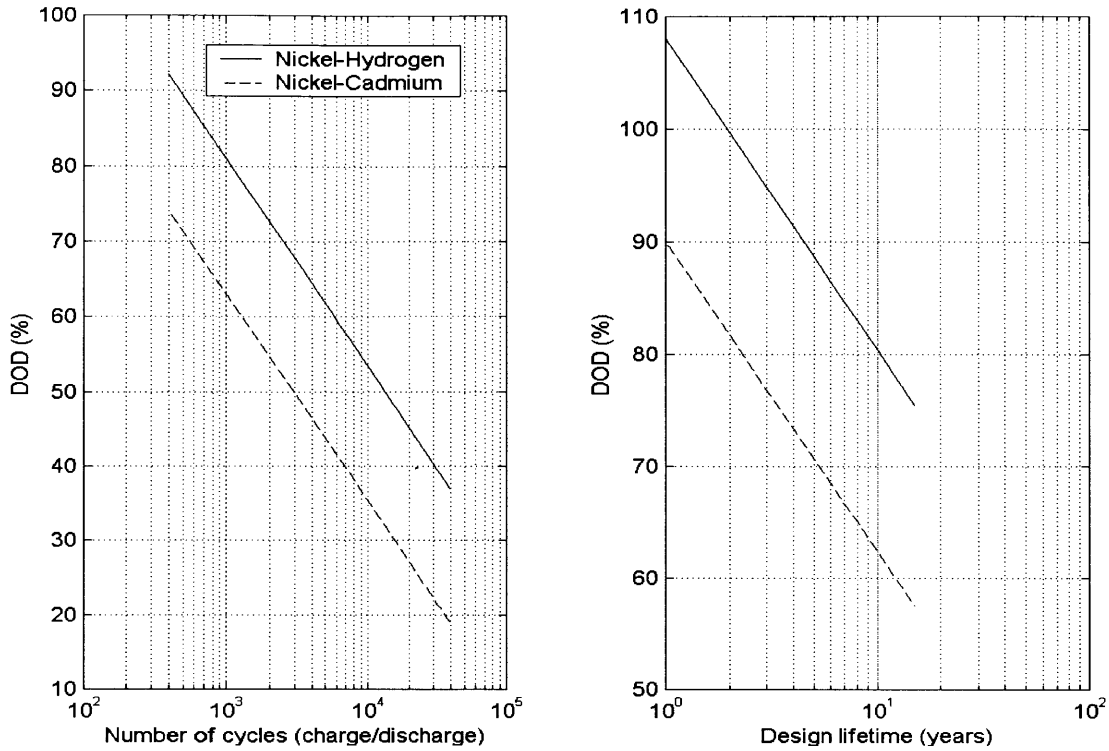
Fig. 4-5. Depth-of-Discharge (DOD) as a function of number of cycles and design lifetime for a GEO satellite.

For instance, for a 3-year mission in GEO, the average DOD for a Nickel-Cadmium battery is approximately 76%, but it drops to 62% for an extended mission of 10 years. How does this impact the sizing of the battery? Battery capacity is estimated as follows:

$$C_r = \frac{P_e \times T_e}{(DOD) \times N \times n} \tag{4-4}$$

P_e = Power requirement during eclipse (W), T_e = Duration of eclipse (hours), N = Number of batteries, n = Transmission efficiency between batteries and load, typically 90%.

The battery capacity scales inversely with the DOD. Therefore, as the number of cycles or the design lifetime increases, the energy available from the batteries during each cycle decreases, i.e., the DOD decreases. Consequently, the batteries have to be over-designed as the design lifetime increases. The mass of batteries can be obtained given the specific energy density of the battery. For Nickel-Cadmium batteries, the specific energy density ranges between 25-30KWh/kg, and 40-60KWh/kg for Nickel-Hydrogen. Lithium-Ion and Lithium-Carbon are expected to reach the

100KWh/kg level. Figure 4-7 shows the evolution of battery mass as a function of design lifetime. The power delivered during eclipse is maintained constant.
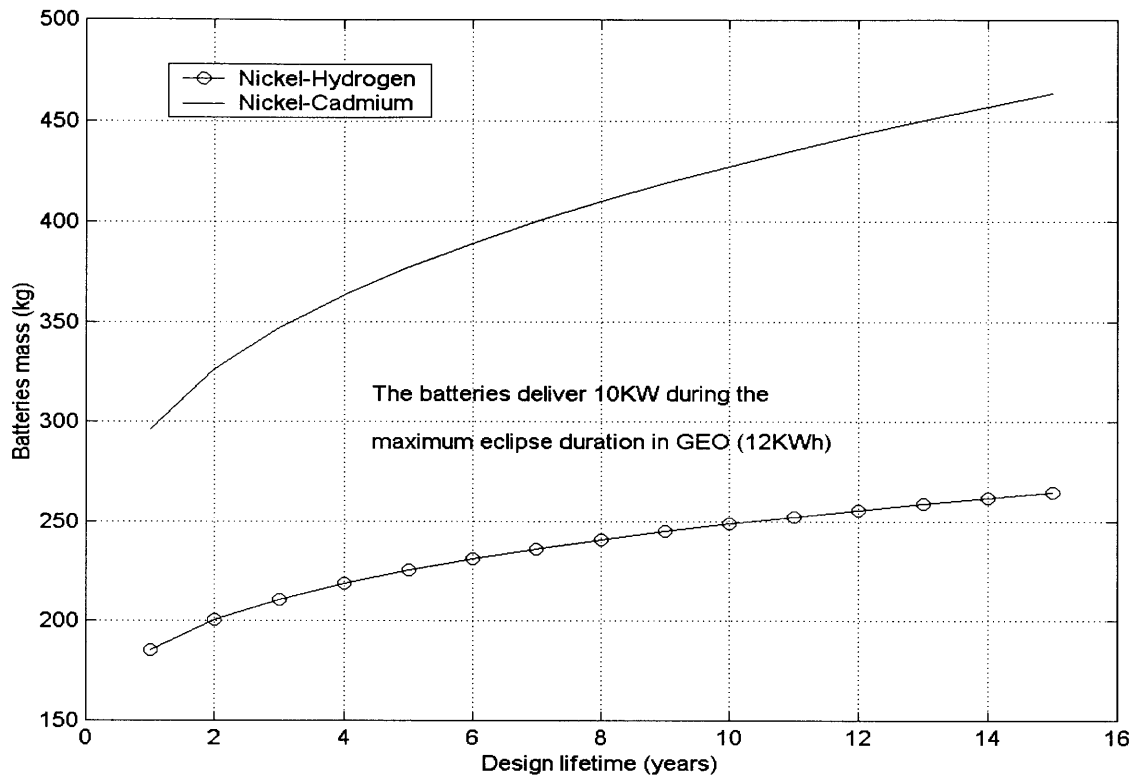
Fig. 4-6. Mass of a battery required to deliver 12KWh as a function of design lifetime.

Figure 6 shows the advantage of Nickel-Hydrogen batteries over Nickel-Cadmium for high-energy capacity requirements and long mission duration. For smaller capacities, the mass of the Nickel-Hydrogen batteries, the mass penalty and the percent mass penalty considering a three years reference mission as a function of the design lifetime are given in Figure 4-7.
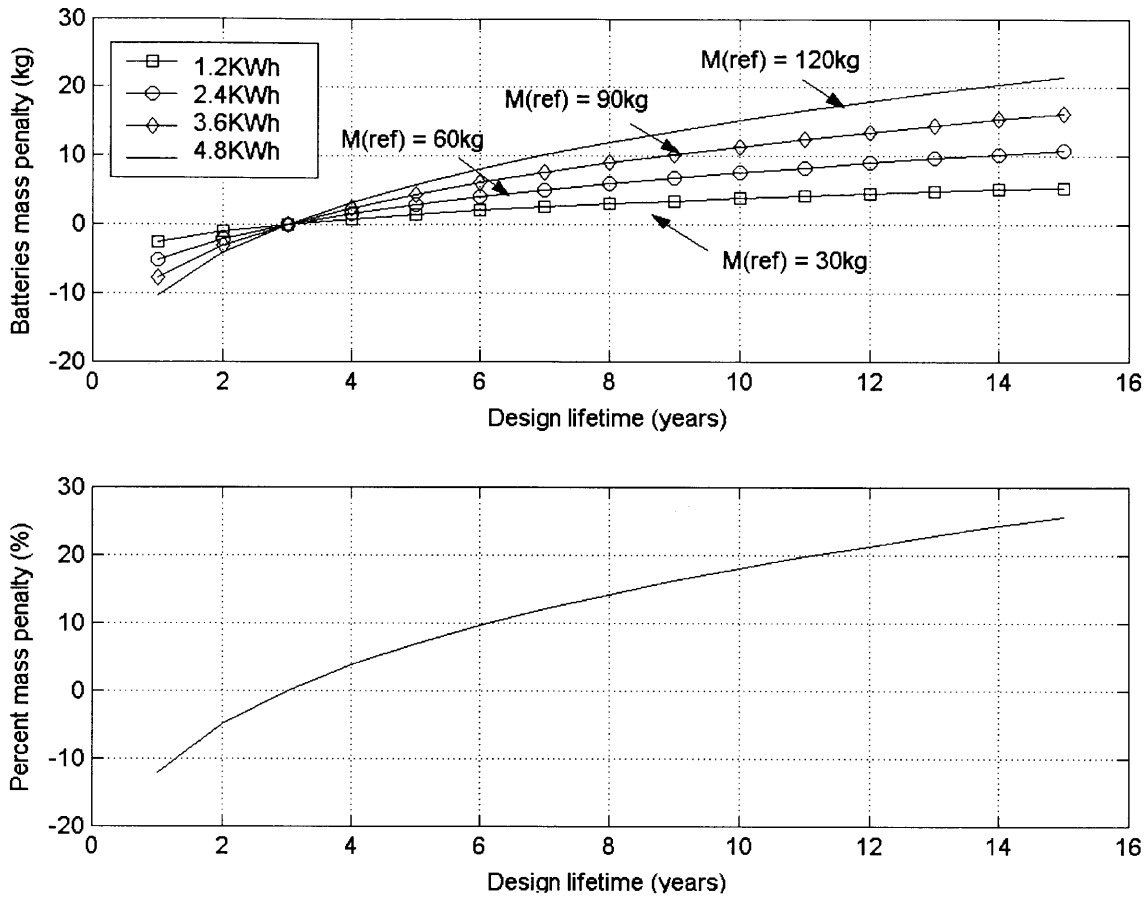
Fig. 4-7. Nickel-Hydrogen batteries, mass, mass penalty, and percent mass penalty as a function of the design lifetime. The reference mission is 3 years.

Power control unit, cables and harnesses: The power distribution system (or sub-subsystem) consists of cabling, fault protection, and switches in the form of mechanical or solid-state relays to turn power on and off to the spacecraft loads. Power regulation is required for two main tasks: 1) Controlling the solar array power output to prevent battery overcharging and spacecraft heating; and, 2) Regulating the spacecraft power bus voltage (or each load separately).

The solar array output is described by a plot of current versus voltage. This I-V curve changes both due to seasonal variation in the array temperature and the solar intensity, and due to radiation degradation of the solar cells as previously discussed . The array voltage is maximum as the spacecraft comes out of eclipse when the temperature of the cells is minimum. Hence the need to regulate the solar array output [Agrawal, 86].

An unregulated bus has a voltage that varies significantly. This is often unacceptable for most of the electronic equipment of the payload and the spacecraft if voltage regulation is not provided separately at each load or equipment. Voltage regulators and converters are therefore either placed separately at each load or on the spacecraft power bus [Wertz and Larson, 99].

It is difficult to quantify how the mass of the power control unit and the power distribution system scale with the design lifetime. The power control unit as well as the cabling and harness are indirectly affected by the design lifetime as excess power is required at beginning-of-life, and increases with design lifetime. We use a mass estimate relationship to evaluate the mass of the power control unit and the power distribution system:

$$M_{PCU}(kg) = 0.0045 \times P_{BOL}(W)$$ (4-5)

The mass of the power distribution system is a large part of the EPS mass, roughly 10% to 20%:

$$M_{dist} = 0.15 \times M_{EPS}$$ (4-6)

Figure 4-8 shows a typical mass breakdown of the EPS for a spacecraft in GEO in terms of its components, solar array, batteries, power control unit, and power distribution, as a function of the design lifetime:

$$M_{EPS} = M_{array} + M_{batteries} + M_{PCU} + M_{dist}$$ (4-7)
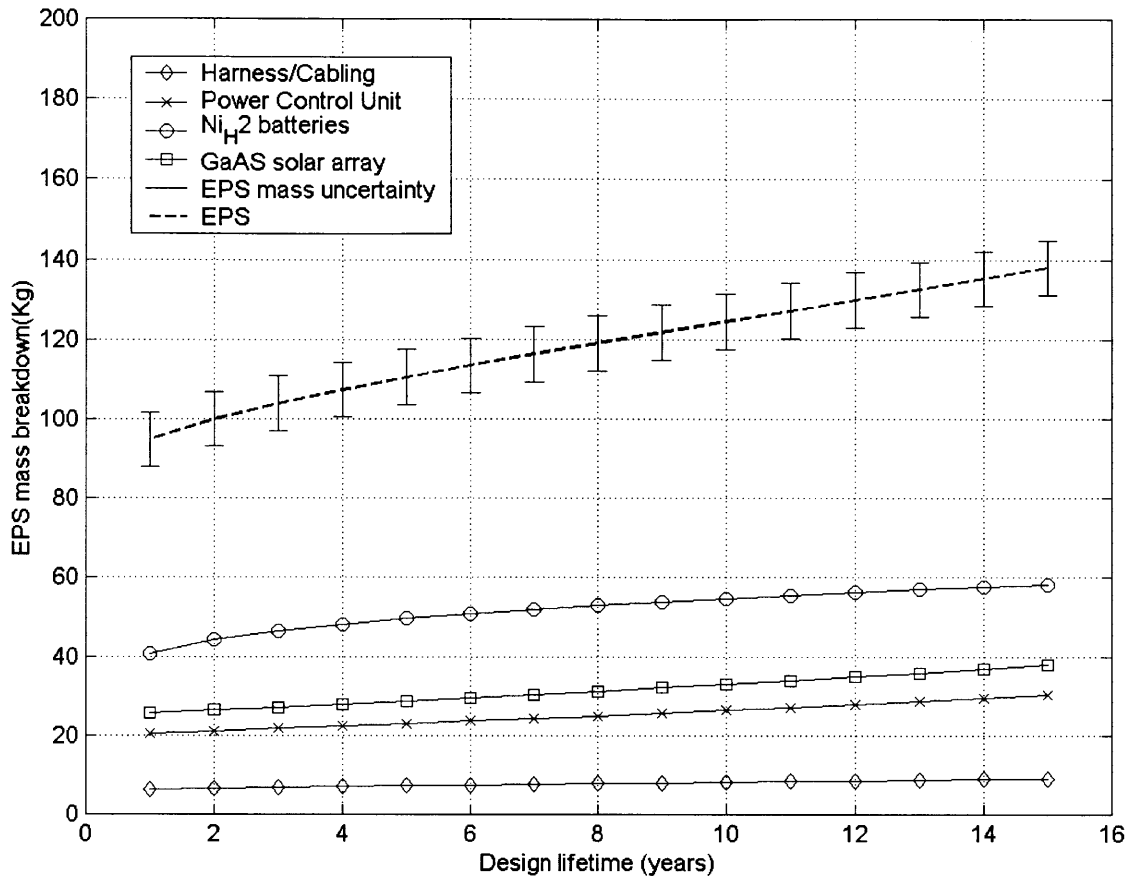
Fig. 4-8. EPS mass breakdown as a function of the design lifetime for a 1KW EOL power requirement.

The mass, mass penalty, and percent mass penalty for the electrical power subsystem as a function of the design lifetime are given in Figure 4-9. The design lifetime for the reference mission is three years.
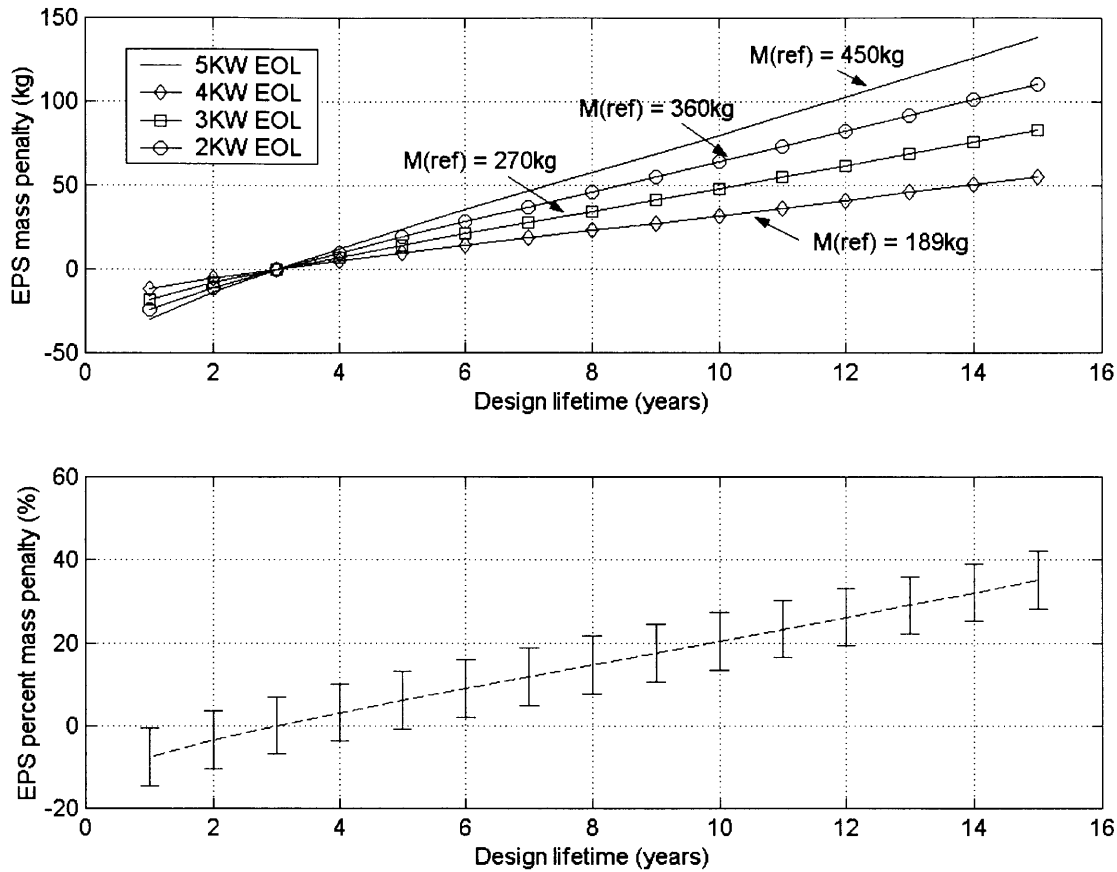
Fig. 4-9: Electrical power subsystem mass, mass penalty, and percent mass penalty as a function of design lifetime (Nickel-Hydrogen batteries and GaAs cells, reference mission = 3 years, satellite is in GEO).

Caveat: The previous sections presented a simple design process for sizing the solar arrays and the batteries. A limited number of parameters were considered, as well as two mass estimate relationships, in order to derive typical mass profiles of the EPS as a function of the design lifetime. These parameters included the power at end-of-life requirement, the spacecraft orbital parameters (to derive the eclipse duration for the sizing of the batteries, and the solar arrays *degradation-per-year*), the solar cell type or the cell energy conversion efficiency, the array specific performance (W/m² or W/kg), and the battery type or its specific energy density. The purpose of this analysis was to highlight and capture the impact of the design lifetime on the sizing of the EPS in a semi-quantitative way. In reality, the design process of the solar arrays and the batteries is much more involved: designers have a plethora of variables to trade and optimize. More elaborate design processes of the EPS are available in the literature [Wertz and Larson, 99] and [Agrawal, 86].

4.3.2 Thermal Subsystem

A spacecraft contains many components that will function properly only if they are maintained within specified temperature ranges. The thermal design of a spacecraft involves identifying the sources of heat, designing proper heat transfer between all spacecraft elements, and rejecting heat so that different components stay within their operating temperature ranges [Wertz and Larson, 99].

As in the previous sections, we are interested in how the thermal subsystem's mass scales with the design lifetime. It is useful to keep in mind in the following discussion that the thermal subsystem accounts on average for only 6% of spacecraft's dry mass (see Table 4-1). A spacecraft's thermal design is highly dependent on the mission class and the attitude stabilization type. Assuming a configuration of the thermal subsystem has been selected for a reference mission (selection of a passive versus active thermal control, thermal coating and multi-layer insulation, heat pipes, louvers, radiators, electrical heaters, etc.), should the subsystem be redesigned if the spacecraft design lifetime varies? If so, how does its mass scale with the design lifetime?

In order to answer the above questions, we first need to look into the different sources of heat that affect a spacecraft. These include solar radiation, Earth albedo and infrared radiation, and equipment power dissipation (electrical components and wiring). While the first two are not affected by the design lifetime, it was shown previously that the power requirement at beginning-of-life increases as the design lifetime increases due to solar array degradation. This excess power (see Fig. 4-2) must be handled by the thermal subsystem. It is therefore reasonable to assume that the thermal subsystem additional mass varies as a function of the difference between P_{BOL} and P_{EOL}:

$$\Delta M_{thermal} = f\left(P_{BOL} - P_{EOL}, T_{Life}, ...\right) \qquad (4\text{-}8)$$

Radiators are sized for the hottest conditions. The heat-balance equation can be written as follows:

$$\varepsilon \sigma T_{rad-max}^4 \eta_r A = \alpha A I_s \sin(\theta) + P \qquad (4\text{-}9)$$

The following notation is used in Eq. 4-9: A = Area of radiator (m^2), ε = Emittance of radiator, σ = Stefan-Boltzmann constant ($5.68*10^8 W/m^2.K^4$), α = Absorptance of the radiator, T_{rad_max} = Maximum allowable temperature for the radiator (K), θ = Solar aspect angle (rad), η_r = Radiator efficiency, typically 90%.

The area of the radiator, and consequently its mass, is proportional to the power dissipation:

$$A = \frac{P}{\varepsilon\sigma T^4\eta - \alpha I_s \sin(\theta)} \tag{4-10}$$

Another effect has to be considered in sizing the radiator surface: the degradation of the thermal properties of its surface. Typically for optical solar reflector (OSR) covering the radiator panel of a spacecraft in GEO, the solar absorptance and emittance vary as shown in Table 4-3.

Table 4-3. Thermal properties of OSR at BOL and EOL after 7 years in GEO. Adapted from [Agrawal, 86].

	Solar Absorptance, α	Solar Emittance, ε
BOL	0.08	0.85
EOL	0.21	0.85

The radiator area has to be sized for the worst case. Assuming that the power dissipation is a fraction of the electric power delivered by the solar panels, the radiator's surface can be estimated as follow:

$$A = \max\left[\frac{k \times P_{BOL}}{\varepsilon_{BOL}\sigma T^4\eta - \alpha_{BOL}I_s \sin(\theta)}; \frac{k \times P_{EOL}}{\varepsilon_{EOL}\sigma T^4\eta - \alpha_{EOL}I_s \sin(\theta)}\right] \tag{4-11}$$

It is not clear which term dominates in the above relationship. The variations of the solar absorptance $\alpha(t)$ and emittance $\varepsilon(t)$ as a function of time depend on the several parameters such as the surface material and the orbital parameters of the spacecraft. In the above example, the first term drives the sizing of the radiator's area. To first order, we will consider the mass of the radiator to be proportional to P_{BOL}:

$$M_{rad} = k' \times P_{BOL} \tag{4-12}$$

We will also assume that the rest of the thermal subsystem mass scales with the mass of the radiator:

$$M_{thermal} = k^{''} \times P_{BOL}$$

For an active thermal control subsystem, values of $k^{''}$ that reflect realistic thermal control subsystem mass range between 0.020kg/W and 0.035kg/W.
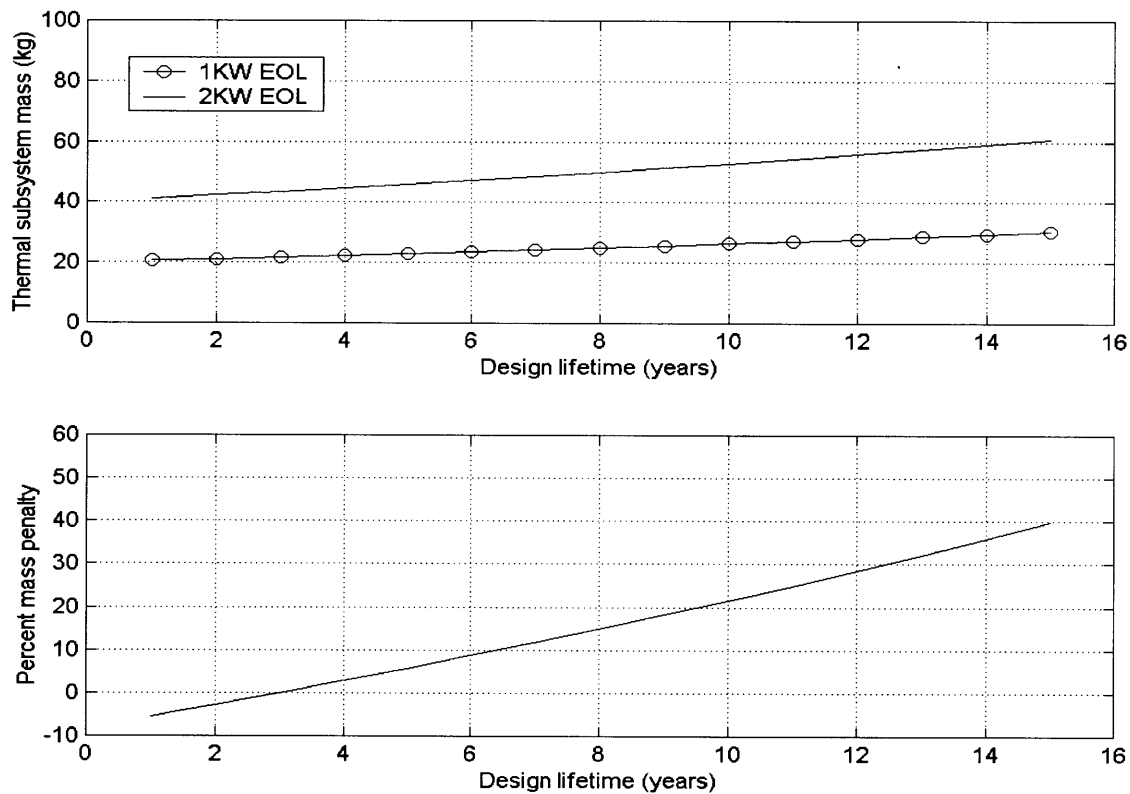
Fig. 4-10. Thermal subsystem mass and mass penalty as a function of design lifetime.

The above discussion represents a first attempt at quantifying the effects of the design lifetime on the thermal subsystem. Although it is clear that the thermal subsystem scales with the design lifetime due to the excess power at BOL and the degradation of the thermal insulation optical properties, it is nevertheless difficult to quantify those effects in a reasonably accurate way without taking into account a multitude of parameters regarding the spacecraft configuration, the type of thermal control, etc., as well as particular details about the mission. Such details are beyond the scope of this study.

94

4.3.3 Telemetry, Tracking, and Control Subsystem

The Telemetry, Tracking, and Control (TT&C) subsystem interfaces between the spacecraft and the ground segment. This subsystem provides the hardware required for the reception, processing, storing, multiplexing, and transmission of satellite telemetry data [Garrison, *et* al., 95]. The Command and Data Handling subsystem (C&DH), often subsumed under the TT&C subsystem, performs two categories of function: it receives, validates, decodes, and distributes commands to other spacecraft subsystems, and gathers, processes, and formats spacecraft housekeeping data for downlink or use by the on-board computer [Wertz and Larson, 99].

As in the previous sections, we are interested in how the TT&C and the CD&H subsystems' mass scale with the design lifetime. Table 4-1 shows that that those two subsystems account on average for 5% of a satellite's dry mass. As with the thermal subsystem, these are minor contributors to the spacecraft mass. The TT&C design is driven by the following requirements: Data rates for command and telemetry, data volume and storage type, uplink and downlink frequencies, bandwidths, receive and transmit power, beamwidth, and antenna characteristics. Selection criteria for TT&C include performance (BER, noise figure, etc.), compatibility with other existing systems (e.g., TDRSS), as well as technology risk [Wertz and Larson, 99].

These requirements as well as the selection criteria do not depend on the spacecraft design lifetime. It is therefore reasonable to assume that, to a first order, the mass of the TT&C does not depend on the design lifetime requirement. The same is true for the C&DH subsystem.

This argument breaks down however if we consider the effect of radiation on the on-board electronics and the need to provide additional shielding as the design lifetime increases. Furthermore, the reliability required of the C&DH will affect the subsystem mass as the design lifetime increases: Redundant components will be needed in order to maintain the same level of reliability for an extended lifetime, hence increasing the amount of hardware and consequently the mass of the subsystem.

We will consider that the mass of the Command and Data Handling as well as the Telemetry, Tracking and Control (C&DH/TT&C) subsystems scale with the level of redundancy n. This approach is further elaborated in the following section.

4.3.4 Reliability/Redundancy issues

This section is, to a large extent, based on section 19.2 in [Wertz and Larson, 99], *Reliability for Space Mission Planning* by H. Hecht, and P. Babcock's *Introduction to Reliability and Modeling of Fault-Tolerant systems* [Babcock, *undated*].

The question we seek to answer or gain insight into is: How does the spacecraft mass scale with design lifetime and mission reliability? Design lifetime is the intended operational time of the spacecraft on-orbit. *Mission reliability* is defined as the probability that the space system will function without a failure that impairs the mission, over a specified period of time or amount of usage. The elementary expression for the reliability of a single product is:

$$R = e^{-\lambda t} \tag{4-13}$$

For a spacecraft composed of n non-redundant elements all equally essential to the spacecraft operation, the overall series reliability is:

$$R_s = \prod_1^n R_i = e^{-\sum_i \lambda_i t} \tag{4-14}$$

λ_i is the failure rate of subsystem I, and R_i is Probability that subsystem i is operational. For n parallel or redundant elements, the overall parallel reliability is:

$$R_p = 1 - \prod_1^n (1 - R_i) \qquad (4\text{-}15)$$

When the reliability of the elements is the same, the above equation simplifies to:

$$R_p = 1 - (1 - R)^n \qquad (4\text{-}16)$$

Consider a reference design lifetime T_{ref} along with a reference mission reliability R_{ref}. As the design lifetime increases, the mission reliability decreases. In order to maintain the same reliability R_{ref} for an extended duration $T_{life} > T_{ref}$, n redundant elements should be considered:

$$n = \frac{\log(1 - R_{ref})}{\log\left(1 - R_{ref}^{\frac{T_{life}}{T_{ref}}}\right)} \qquad (4\text{-}17)$$

This *level of redundancy* can be calculated per spacecraft subsystem (assuming same components are selected). Consequently, a spacecraft subsystem's mass will scale with its level of redundancy, assuming the customer/designers want to maintain a reliability at end-of-life, R_{EOL}, constant and independent of design lifetime.

As a simple model, we will assume that the mass of the Command and Data Handling as well as the Telemetry, Tracking and Control (C&DH/TT&C) subsystems scale directly with n.
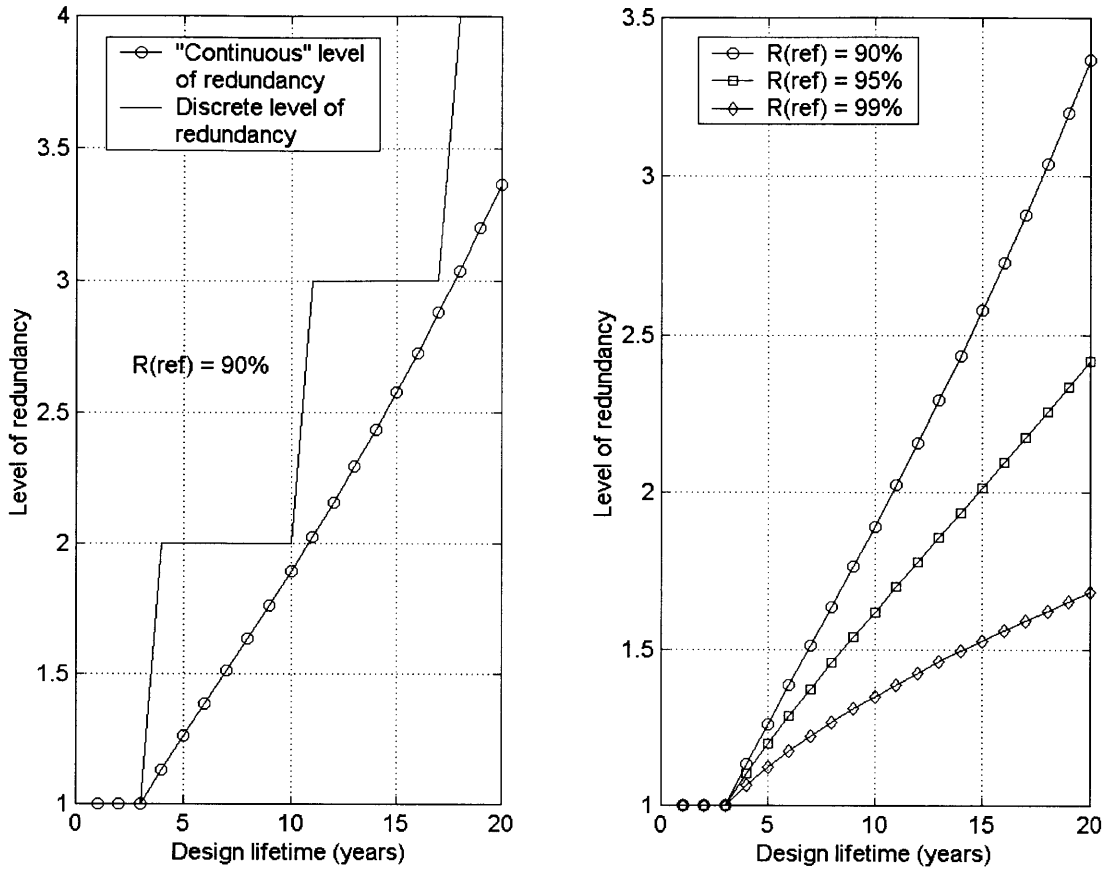
Fig. 4-11. Level of redundancy as a function of the design lifetime. The reference mission is 3 years (mission reliability or subsystem reliability).

4.3.5 Propellant Budget

As in the previous sections, we are interested in how the propellant mass scales with the satellite design lifetime. Propellant is required for orbit change, orbit maintenance, and attitude control. The propellant budget includes propellant to change spacecraft orbital parameters (e.g., orbit transfer), correct for errors due to dispersion injection, control the attitude during thrusting, counter disturbance forces (e.g., drag in LEO or third-body gravitational attraction in GEO), and correct spacecraft angular momentum. It also includes a provision for end-of-life disposal (ΔV to de-orbit if the spacecraft is in LEO, or to raise the altitude if the spacecraft is in GEO), as well as a *propellant margin* that consists of a percentage of the identified propellant requirement.

The total velocity change ΔV_{tot} is converted to propellant mass as follows:

$$M_p = M_0 \left[1 - e^{-\left(\frac{\Delta V}{I_{sp}g} \right)} \right]$$
(4-18)

M_p is the mass of propellant required for a given velocity increment, M_0 is the initial spacecraft mass. As stated previously, the satellite ΔV_{tot} is made up of four parts: ΔV_{init} for initial orbit insertion, ΔV_{yr} for yearly station keeping as well as reaction wheel desaturation or "unloading", ΔV_{EOL} for end-of-life disposal, and a ΔV_{margin} [Lamassoure and Hastings, 01].

$$\Delta V_{tot} = \Delta V_{init} + \Delta V_{yr} \times T_{life} \left(years \right) + \Delta V_{EOL} + \Delta V_{margin}$$
(4-19)

We will focus on the first two components of the propellant budget. Two parameters that vary with the design lifetime affect the propellant budget: The satellite initial mass M_0 and the ΔV_{stk} required for station-keeping over the mission duration.

- For instance, given a ΔV_{init} requirement for orbit transfer from GTO to GEO, the propellant mass needed to provide this velocity increment is a linear function of M_0, as illustrated in Eq. 4-18. Since M_0 varies as a function of the design lifetime T_{life}, so does $M_{p_{init}}$:

$$M_{p_{init}} \left(T_{life} \right) = \psi \left(\Delta V_{init}, I_{sp} \right) \times M_0 \left(T_{life}, \ldots \right)$$
(4-20)

ψ is given in Eq. 4-18. For example, for $\Delta V_{init} = 1500$m/s, and an $I_{sp} = 300$s, $\psi = 0.4$. In other words, the propellant required to perform the orbit transfer accounts for 40% of the spacecraft mass.

- The ΔV_{stk} required for station keeping can be estimated as follows:

$$\Delta V_{stk} = T_{life} \times \Delta V_{yr}$$
(4-21)

The ΔV_{yr} yearly for station keeping is a function of the orbit altitude, the solar cycle (min or max), which in turn alters the atmospheric density, hence the drag encountered by satellite in LEO, or the longitude of station keeping for a satellite in GEO. Typically for a satellite in GEO, $\Delta V_{yr} \sim 50$m/s. Finally, the propellant mass required to provide ΔV_{tot} is given by:

$$M_{P_{tot}} = M_0 \times \left[1 - e^{-\frac{\left(\Delta V_{init} + \Delta V_{yr} \times T_{life} + \Delta V_{EOL} + \Delta V_{margin} \right)}{g \times I_{sp}}} \right]$$ (4-22)

Station keeping is performed using a separate propulsion system from the orbit insertion system. Increasingly ion propulsion or Hall effect thrusters are used for station keeping because of their high specific impulse ($I_{sp} \sim$ 1500-3000m/s). In this case, for a spacecraft in GEO, the mass of propellant required for station keeping per year accounts for 0.2%–0.4% of the spacecraft mass at BOL, as opposed to 1.5%–3% using the more traditional chemical propulsion system.

4.3.6 Propulsion Subsystem

The propulsion module subsystem (PMS) consists of the tanks to hold the propellant, pipes and pressure-regulating equipment, and the thrusters [Wertz and Larson, 99]. As in the previous sections, we are interested in how the propulsion subsystem's mass scales with the design lifetime, keeping in mind that the propulsion subsystem accounts on average for 4% of a spacecraft dry mass (see Table 4-1).

$$M_{propulsion} = M_{tank} + M_{pipes/valves} + M_{thrusters}$$ (4-23)

As the design lifetime increases, the propellant budget increases. Consequently, the volume and mass of the tank necessary to hold the propellant increase. It is reasonable to assume that the other contributors to the propulsion subsystem mass remain unaffected by an increase in the design lifetime.

Assuming a thin spherical tank of thickness e and radius r, the mass of the tank and the mass of its propellant are:

$$\begin{cases} M_{tank} \cong \rho_{tank} \left(4\pi r^2 \right) \times e \\ M_{propellant} = \rho_{propellant} \left(\frac{4}{3} \pi r^3 \right) \end{cases}$$ (4-24)

Consequently

$$M_{tank} = \left(4\pi \rho_{tank} e \right) \times \left(\frac{3}{4\pi \rho_{propellant}} \right)^{\frac{2}{3}} \times \left(M_{propellant} \right)^{\frac{2}{3}}$$ (4-25)

Finally, we can relate the propulsion subsystem's mass to the propellant mass, which varies as a function of the design lifetime, with the following functional relationship:

$$M_{propulsion} = a + b \times M_{propellant}^{\frac{2}{3}} \qquad (4\text{-}26)$$

a and *b* are constants and depend on the particular design of the propulsion subsystem. They do not vary with the design lifetime.
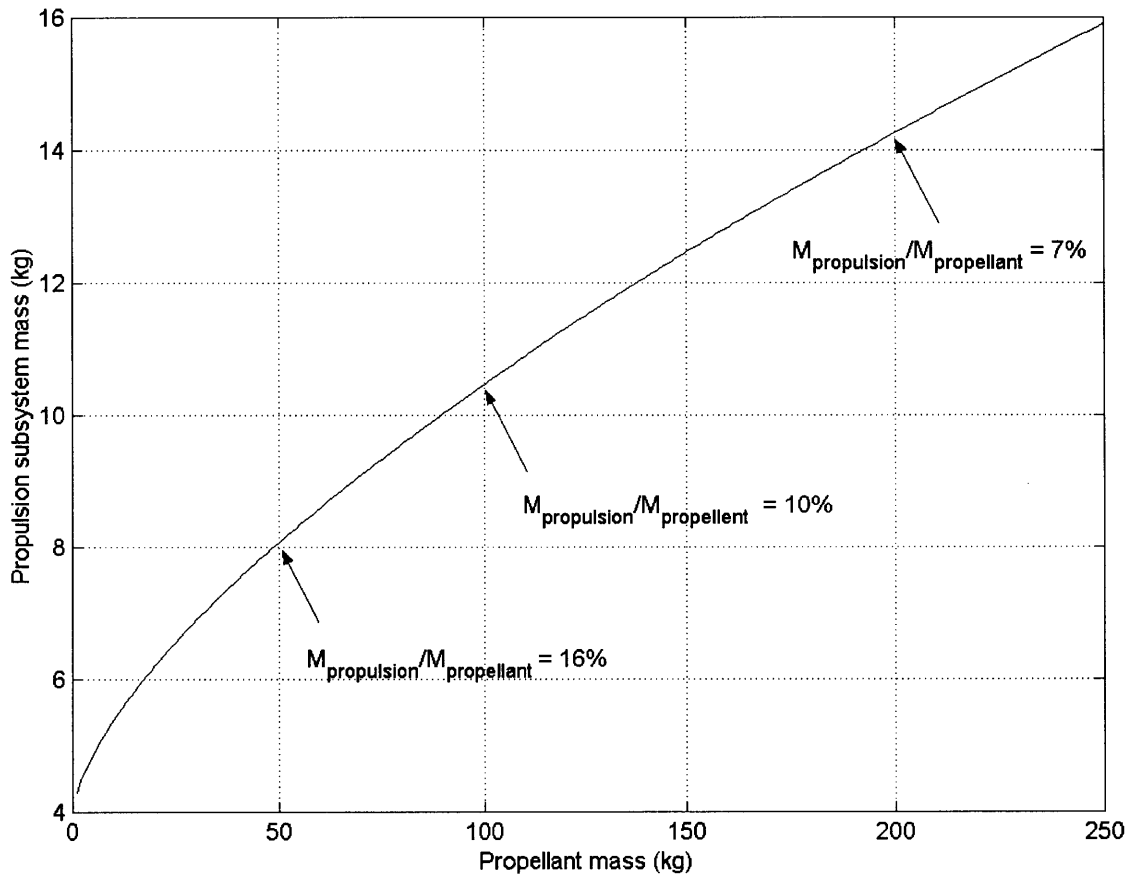
Fig. 4-12. Typical propulsion subsystem mass as a function of propellant mass (a = 4, b = 0.3).

4.3.7 Attitude Determination and Control Subsystem

The ADCS measures and controls the spacecraft's angular orientation. This subsystem stabilizes the spacecraft in desired orientations during different mission phases despite disturbance torques (thruster misalignment, aerodynamic torque, solar radiation torque, etc.), and is also used to re-orient the spacecraft in order to point the payload in different directions (slew maneuvers). Its mass accounts for on average 6% of a satellite dry mass (see Table 4-1).

The issue of concern in this section is how does the ADCS scale with the spacecraft design lifetime. The selection and sizing of the ADCS is driven by requirements on accuracy and range of angular motion both in terms of determination and control. For a three-axis stabilized spacecraft. The torque capability or control authority of reaction and momentum wheels is determined by the magnitude of the disturbance torques, and the elements of the spacecraft inertia matrix.

$$T_{ADCS} = f(T_{dist}, \underline{I}, ...)$$ (4-27)

For a mass M with an orthogonal coordinate system (x, y, z) located at its center of mass, the moment of inertia about the z-axis for instance is given by:

$$I_{zz} = \int_M (x^2 + y^2) dM$$ (4-28)

As the design lifetime increases, the spacecraft mass increases (EPS, thermal subsystem, propellant, etc.), hence the elements of its inertia matrix. Consequently the ADCS has to be redesigned for a larger torque capability. How can we relate the torque capability of a wheel to its mass? Answering this question would provide an insight into the relationship between the ADCS mass and the spacecraft design lifetime. This step unfortunately is not straightforward. In the absence of a physical based rationale for relating the ADCS mass to the design lifetime, we will use as a substitute the mass estimate relationship provided in Table 4-1 to evaluate the mass of a three-axis ADCS.

$$M_{ADCS} = 0.06 \times M_{dry}$$ (4-29)

4.3.8 Structures

The function of a spacecraft structure is to provide mechanical support to all subsystems within the framework of the spacecraft configuration. It also satisfies the subsystem requirements, such as alignment of sensors, actuators, antennas, etc.; and the system requirements for launch vehicle interfaces and integration [Agrawal, 86]. The spacecraft structure is a major contributor to the spacecraft dry mass and accounts for 21% of its dry mass (see Table 4-1). As in the previous sections, we are interested in how the spacecraft structure mass scales with the design lifetime. In order to address this question, we start by examining the sources of structural requirements. Structures must endure mechanical loads in different environments, from manufacturing, to launch and normal operations [Wertz and Larson, 99]. The environments from which the structural requirements are derived are listed in the Table 4-4.

Table 4-4. Sources of structural requirements by mission phase. Adapted from [Wertz and Larson, 99]

Environment/Phase	Source of Requirements
Manufacturing and assembly	Handling fixtures, stresses induced by welding, etc.
Transport and handling	Crane or dolly reactions, land, sea, air transport environments
Testing	Vibrations and acoustic tests, test fixtures reaction loads
Pre-launch	Handling during stacking sequence, pre-flight checks
Launch	Steady-state booster acceleration, acoustic noise, transient loads during booster ignition, burn-out, pyrotechnic shock from separation events

None of the above items can clearly relate the spacecraft design lifetime to the structural requirements, and consequently to the spacecraft structure mass. It is reasonable to assume that the spacecraft structure scales with the design lifetime by the fact the different subsystems *enclosed* within or supported by the structure, as well as the consumables, scale with the design lifetime (EPS, thermal, propulsion, propellant). It is not obvious, however, how the structure mass scales with the design lifetime. The least arbitrary approach is to maintain the mass estimate relationship given in Table 4-1 relating the spacecraft structure to the satellite dry mass

$$M_{struct} = 0.21 \times M_{dry} \qquad (4\text{-}30)$$

4.4 Spacecraft Mass Profile

The spacecraft mass profile as a function of the design lifetime can now be sketched by combining the effects of the design lifetime on the different subsystems as discussed in the previous sections. The independent variables include:

- Orbit type and related parameters (eclipse duration, number of batteries charge/discharge cycles, *degradation-per-year* of solar arrays, ΔV_{yr})
- Solar cells type and battery type
- Power at end-of-life
- Mission reliability
- Type of attitude control
- Payload mass

The spacecraft dry mass and total mass (loaded mass) are calculated as follows:

$$
\begin{cases}
M_{dry} = M_{EPS} + M_{thermal} + n \times M_{TT\&C+CD\&H} + M_{ADCS} + M_{propulsion} + M_{struct} + M_{payload} \\
\\
M_{tot} = M_{dry} + M_{propellant}
\end{cases}
\tag{4-31}
$$
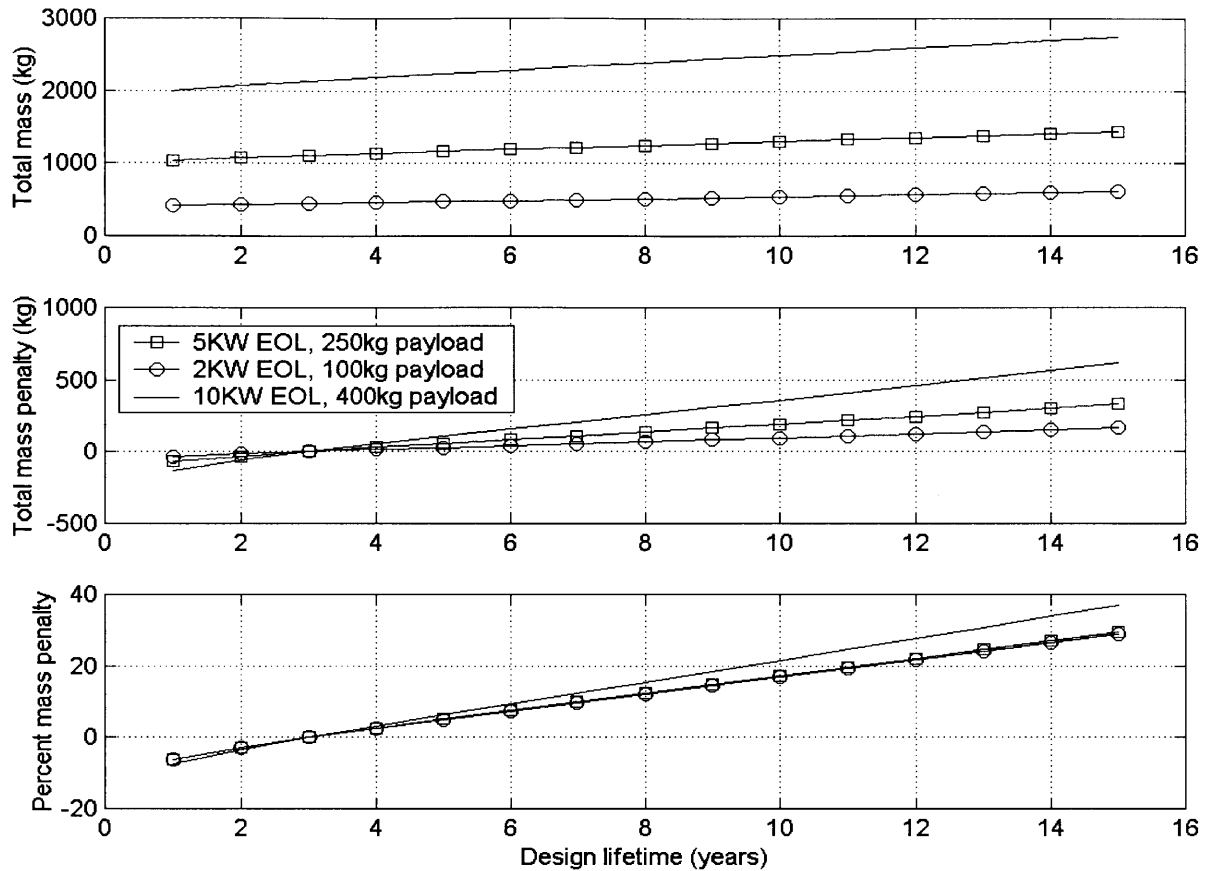
Fig. 4-13. Spacecraft total mass, mass penalty, and percent mass penalty as a function of the design lifetime (spacecraft in GEO, mission reliability = 95%, three-axis stabilized, GaAs cells, Ni-H₂ batteries, reference mission = 3 years).

Figure 4-13 shows typical spacecraft mass profile, mass penalty, and percent mass penalty as a function of the design lifetime. It is interesting to note for instance that designing a spacecraft for 3 years instead of 15 years results in a mass saving of the order of 40%. Or conversely, a mass penalty of 40% is incurred if a mission is initially designed for 15 years instead of 3 years. The next step is to translate this mass penalty, or mass saving, into a cost penalty, or cost saving. This is undertaken in the next section.

4.5 Cost to IOC and Cost-per-Operational Day

This section is based to a large extent on Chapter 20 of *Space Mission Analysis and Design* [Wertz and Larson, 99] by H. Apgar, D. Bearden, and R. Wong, as well as on Chapter 8 of *Reducing of Space Mission Cost*, by D. Bearden, R. Boudreault, and J. Wertz [Wertz and Larson, 96].

In this section, we are interested in isolating the effect of the design lifetime on the spacecraft cost. We will proceed by translating the various mass profiles established previously into spacecraft cost profiles as a function of the design lifetime. In order to do so, an understanding of the rationale, advantages and limitations of Cost Estimate Relationships, as well as the various components of a spacecraft cost is required. The following paragraphs summarize the basics of cost modeling.

A spacecraft's cost depends on its size, complexity, technology readiness (TRL), design lifetime, schedule, as well as other characteristics. Space systems have specific costs (cost-per-unit weight) of the order of $70,000 per kilogram [Wertz and Larson, 96]. Specific costs however are not sufficient for predicting *real* costs of spacecraft. Over the years, several governmental organizations have developed Cost Estimate Relationships (CERs) that relate spacecraft cost or subsystem cost to physical, technical, and performance parameters. The CERs are based on an appropriate historical database of past satellite programs. The basic assumption of parametric cost modeling is that "satellites will cost next time what they cost the last time." CERs include both non-recurring and recurring costs associated with a space system. Non-recurring costs are commonly referred to as the Research, Development, Test, and Evaluation (RDT&E) costs. These costs include the design, analysis and test of prototypes and qualification units. Recurring costs include the cost to produce flight units. They are commonly referred to as the Theoretical First Unit (TFU) costs. This concept represents the cost of the first space-qualified satellite. Typical CERs include the range of the parameters used to develop the correlations between the subsystems characteristics and their cost, the CER itself, and the associated standard error (SE). An example is given in Table 4-5.

Table 4-5. CER for estimating subsystems TFU and RDT&E costs. Adapted from [Wertz and Larson, 99]

Component	Parameter (x)	Range	TFU (FY00$K)	SE (%)	RDT&E (FY00$K)	SE (%)
Structure	weight (kg)	[54kg, 560kg]	13.1x	36	$157x^{0.83}$	38
EPS	weight (kg)	[31kg, 573kg]	$112x^{0.763}$	44	62.7x	57

Launch costs are on the other hand derived from published look-up tables. The *International Guide to Space Launch Systems* [Isakowitz, 99] is the reference for launch systems characteristics and costs. Another approach to modeling launch cost is to evaluate an average cost-per-kilogram

to orbit. For instance, the average cost to LEO per kilogram for both U.S. and European launchers is approximately $K10. Finally, the cost to IOC is given by:

$$Cost_{IOC} = TFU + RDT \& E + Cost_{launch} \qquad (4\text{-}32)$$

Using the linear extrapolation of the launch cost, the cost to IOC can be plotted as a function of the spacecraft design lifetime.
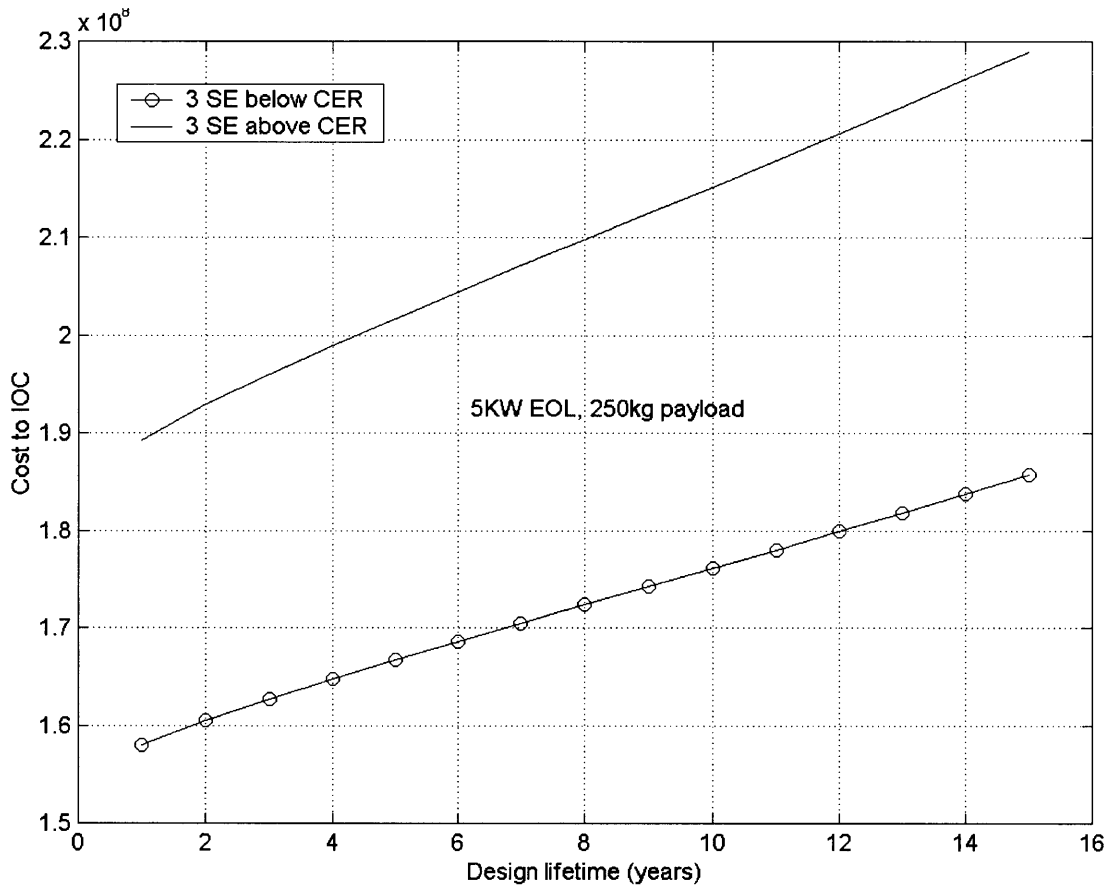
Fig. 4- 14. Cost to IOC for a LEO spacecraft as a function of the design lifetime (three Standard Errors above and below the nominal CER output, same parameters as in Fig. 4-13).

Figure 4-14 shows the range of uncertainty in the cost estimation of the spacecraft recurring and non-recurring costs.
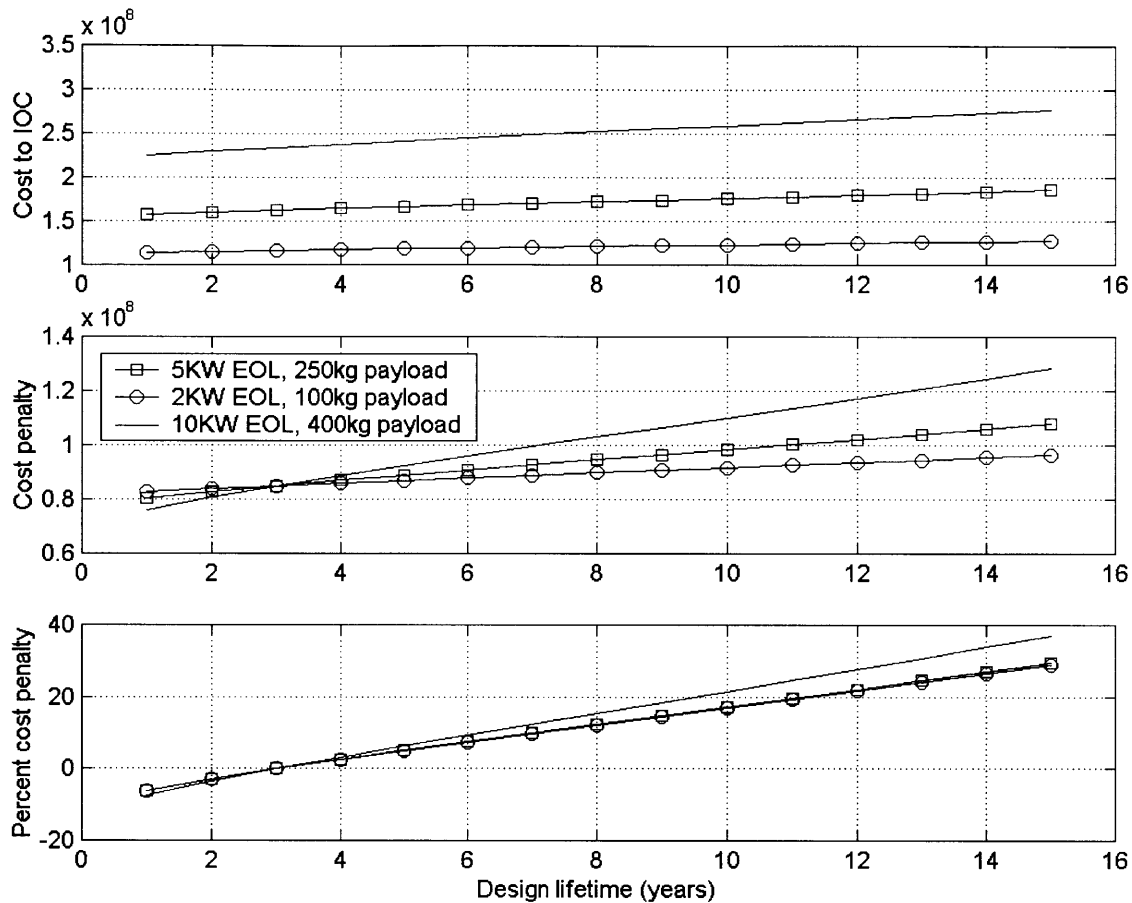
Fig. 4-15. Cost to IOC, cost penalty, and percent cost penalty as a function of the design lifetime (same parameters as in Fig. 4-13)

Figure 4-15 shows typical spacecraft cost to IOC, cost penalty, and percent cost penalty as a function of the design lifetime. It is interesting to note for instance that designing a spacecraft for 3 years instead of 15 years results in a cost saving of the order of 35%. Or conversely, a cost penalty of 35% is incurred if a mission is initially designed for 15 years instead of 3 years.

Figure 4-15 provides an answer to the question we set to investigate in this paper, namely how does the design lifetime requirement, impact the total system (mass and) cost to IOC? The results confirm that the design lifetime is indeed a key driver of the space system cost, and illustrate its particular impact on the various subsystems (EPS, thermal, propulsion, etc.).We can now define the cost-per-operational day of a spacecraft as follows:

$$Cost_{/day} = \frac{Cost\ to\ IOC}{Design\ lifetime\ (days)} \tag{4-33}$$

This metric corresponds to uniformly amortizing the cost to IOC over the entire intended mission duration (without accounting for the time value of money).
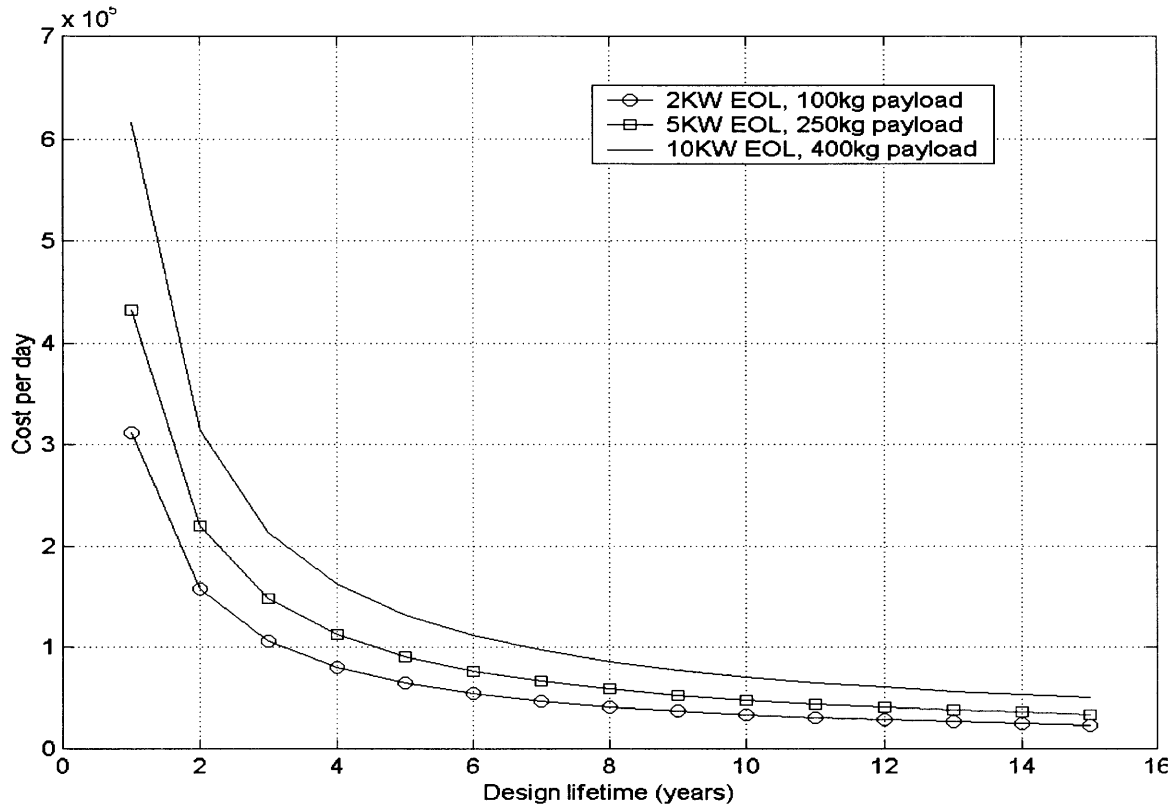
Fig. 4-16. Cost-per-operational day as a function of the design lifetime (same parameters as in Fig. 4-13).

Within the interval of the design lifetime considered, the cost-per-operational day decreases monotonically. In the absence of other metrics, the cost-per-operational day justifies pushing the boundary of the design lifetime and designing spacecraft with increasingly longer lifetimes. It also suggests that a customer is always better off requesting the contractor to provide the maximum design lifetime

$$T_{Life} = T_{Life\_max} \tag{4-34}$$

This, however, is not necessarily true. Launching spacecraft with increasingly longer design lifetimes raises the risk for the satellite of becoming technically and commercially obsolete before

the end of its mission. Thus, in specifying the design lifetime requirement, decision-makers have to assess this risk of loss of value due to both the obsolescence of their product's technology base, as well as the likelihood of changing market needs—or the volatility of the market the system is serving—during the system's operational lifetime. These issues will be explored in the following Chapter.

4.6 Limitations

The analysis discussed above presents several limitations that degrade the accuracy of the results. First, in order to isolate and capture the effects of the design lifetime on the spacecraft mass and cost, a limited number of parameters were considered in the analysis, instead of the plethora of variables that subsystems experts typically have to trade and optimize. This was done is order to maintain a manageable size analysis, and to avoid drowning the key parameters and effects in background clutter.

The second limitation results from the use of mass estimate relationships, such as in the case of the spacecraft structure. While it is clear that the spacecraft structure for instance scales with the design lifetime by the fact the different subsystems enclosed within or supported by the structure scale with the design lifetime (EPS, thermal, propellant, etc.), it is not possible to relate the spacecraft structure's mass to the design lifetime without taking into account particular details about the mission, or the spacecraft configuration and lay-out. In other words, a preliminary design of the spacecraft is required in order to reasonably estimate the mass of the spacecraft structure. In the light of the objectives of this paper, set forth in the introduction and summarized above, such an analysis is beyond the scope of this study. In the absence of quantifiable physical arguments for relating a subsystem's mass to the design lifetime, mass estimate relationships were used as the least arbitrary way to proceed with the analysis.

The third limitation is due in part to the use of cost estimate relationships and dollars-per-pound to estimate launch costs. This resulted in smooth or continuous cost profiles instead of discontinuous profiles that would be obtained in reality because of the performance and cost of existing launch systems (e.g., $13M for less than 1000lb to LEO on Pegasus XL, and $22M for less than 3000lb to LEO on Taurus). The availability and use of commercial-off-the-shelf (COTS) hardware, which exists in discrete *performance bins* and does not necessarily match the customer's needs exactly, will also render discontinuous both the mass and cost profile of a spacecraft as a function of the design lifetime.

Some of the limitations discussed above render the task of building generic models relating the spacecraft mass and cost to the design lifetime very challenging. However, in practice, the above-mentioned inaccuracies will be attenuated when, during the conceptual design phases of a particular spacecraft, designers evaluate the mass and cost of their particular design at discrete values of the design lifetime (e.g., 3, 5, 7, 9 years, etc.), all else being equal (performance and reliability). Thus, more accurate estimates could be obtained for the mass and cost of the spacecraft, or its cost-per-operational day, and help guide the selection of the design lifetime.

4.7 Summary

This Chapter explored the impacts of the design lifetime on the spacecraft mass and cost to IOC. It first examined how different subsystems scale with the design lifetime, using physically based arguments whenever possible, and mass estimate relationships in other instances. The data was then transformed to generate spacecraft mass and cost profiles as a function of the design lifetime. Preliminary results confirm that the design lifetime is a key requirement in sizing various subsystems. For instance, a mass and cost penalty of 30% to 40% is typically incurred when designing a spacecraft for fifteen years instead of three years, all else being equal. It was also shown that the cost-per-operational day decreases monotonically with the design lifetime. This finding justifies pushing the boundary of the design lifetime and designing spacecraft with increasingly longer lifetimes. It also suggests that a customer is always better off asking the contractor to provide the maximum design lifetime achievable. This however may not always be the case. The decision regarding the design lifetime requirement should incorporate external factors such as the obsolescence of the technology embedded in the spacecraft, the relationship between technology obsolescence and market share, and the volatility of the market the mission is serving in the case of a commercial satellite. These issues are discussed in the next Chapter.

Chapter 5

Weaving Time into System Architecture:
Satellite Cost-per-Operational Day and Optimal Design Lifetime

"One should beware of mathematicians and all who make empty prophecies.
The danger already exists that the mathematicians have made a covenant
with the devil to darken the spirit and confine man in the bonds of Hell"

St Augustine, Bishop of Hippo, circa 400 A.D.

5.1 Introduction

What drives a product's design lifetime? How do designers, managers, and customers decide on a system's lifetime requirement, and what is the rationale for specifying this requirement? As the previous chapter highlighted, questions regarding the design lifetime requirement of complex engineering systems can be grouped into three categories:

1. What limits the design lifetime? How far can designers push the system's design lifetime? What is the lifetime "boundary" and why can't it be extended?

2. How do the different subsystems scale with the design lifetime requirement, and what is the total system cost profile as a function of this requirement?

3. What does (or should) the customer ask the contractor to provide for a design lifetime, and why?

The previous chapter investigated the second category of questions for the particular case of space systems. It first examined how different spacecraft subsystems scale with the design lifetime, then integrated these results in order to produce total system mass and cost (to IOC) profiles as a function of the design lifetime. Addressing these issues associated with the second category of questions is pivotal for understanding the rationale in specifying the lifetime requirement. In other words, it is a prerequisite for answering the third category of questions

regarding the specification of the design lifetime. In this chapter, we explore the rationale in specifying spacecraft design lifetime (i.e., the third category of questions stated above regarding the design lifetime). According to Wertz and Larson (1999), the design lifetime requirement, in the case of satellite systems, is "assigned rather arbitrarily" with an understanding of the technical limitations and an intuition regarding the economical impacts associated with designing for longer lifetimes. But what are the economic impacts associated with a system's design lifetime? Can we formally capture them and quantify them? Is there an optimal design lifetime for a satellite that maximizes some economic metric? What characteristics of the system's environment, if any, should be taken into account, in order to select an optimal design lifetime? These are some of the questions that will be addressed in this chapter. A formal process for specifying spacecraft design lifetime is formulated and presented.

This chapter is organized as follows: Section 2 presents the ambiguity regarding a product design lifetime. Section 3 recalls the main results from the previous Chapter regarding the cost profile of a spacecraft as a function of the design lifetime and its cost per operational day. Section 4 introduces the fundamental equation defining the expected present value of a system architecture as a function of its design lifetime. Section 5 explores the existence and the dynamics of the optimal design lifetime under various assumptions (constant revenues per day, technology obsolescence, and market volatility). Section 6 concludes the findings and prepares the way for the following chapter, Flexibility and On-Orbit Servicing.

5.2 To Reduce or to Extend the Design Lifetime? A Conflicted Attitude Towards Products' Design Lifetime

The attitude towards products' design lifetime has often been ambiguous, and at times uninformed. Although issues related to systems design lifetime have received little attention in the literature, there have been qualitative arguments fraught with subjectivity for or against extending a system's design lifetime.

On the one hand, there is a popular belief that manufacturers of durable goods (e.g., automobile tires, light bulbs, batteries) often deliberately reduce the time period that a product remains operational in order to increase their sales and profits. For instance, its seems that the electric lamp industry in the United States, a highly concentrated industry, "has served to limit, and frequently reduce, lamp life in order to increase sales [...] It appears [however] that consumers' interests would generally be better served by a bulb of much longer life" [Avinger, 75]. This

113

hypothetical practice has sparked environmental concerns among ecologists and policy makers, and created interest in the contribution that extended product design lifetime can make towards reducing the waste management and other environmental problems [OECD, 1982]. As a result, a marked stigma has been attached to products designed for short lifetime and the manufacturers that field such designs. Several industries however strongly denied having a concealed policy of accelerated product obsolescence, i.e., deliberately introducing upgrades or new functionalities in a product in order to promote consumer dissatisfaction with existing products and promote sales of new products [Conn, 1978].

On the other hand, in recent years, manufacturers of high-value assets (e.g., rotorcraft, spacecraft) have chosen to increase their products' design lifetimes. Over the last two decades, telecommunications satellites for instance have seen their design lifetime on average increase from seven to fifteen years. In this case, increasing the space segment design lifetime was driven by the desire to maximize the return on investment, and perhaps an intuition that the cost of the system per unit operational time would decrease. However, extending satellite design lifetime has several side effects. On the one hand, it leads to larger and heavier satellites as a result of several factors such as additional propellant for orbit and station-keeping, increased power generation and storage capability, which in turn increases the satellite's development and production cost. On the other hand, as the design lifetime increases, the risk that the satellite becomes obsolete, technically and commercially, before the end of its lifetime increases. This trade-off is illustrated in Figure 5-1.
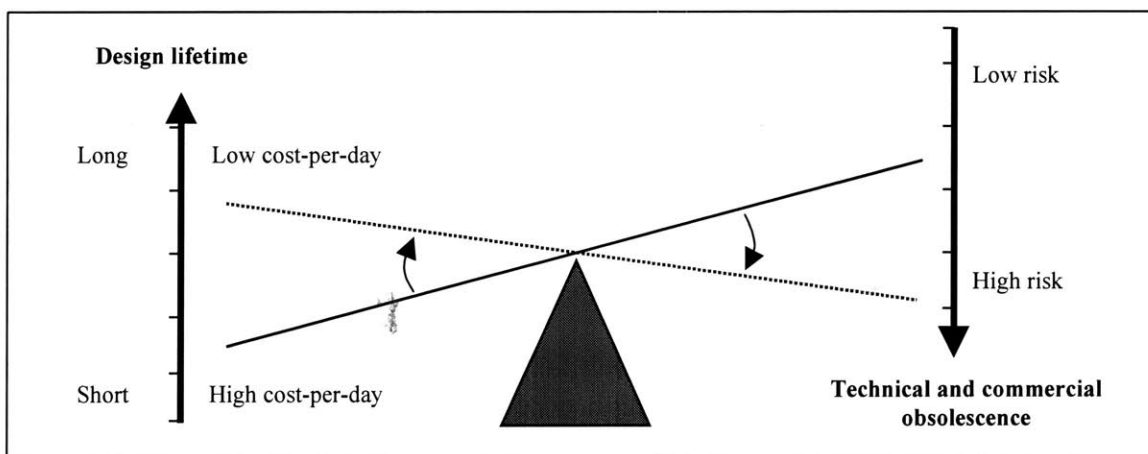
Fig. 5-1. Graphical illustration of the design lifetime trade-off: Extending a satellite design lifetime decreases its cost per operational day. However, it increases the risk that the satellite becomes technically and commercially obsolete before the end of its lifetime.

The discussion above indicates that in specifying the design lifetime requirement, decision-makers have to assess the risk of loss of value due to both obsolescence of their product's technology base, as well as the likelihood of changing market needs after the system has been fielded (volatility of the market the system is serving). For example it is not obvious to be in the best interest of a customer to make the contract life of a spacecraft too long: New or enhanced capabilities, e.g., better spatial resolution for an optical instrument, might be developed and become available within a couple of years following the launch, hence the need to launch a new satellite or risk losing market share to a competitor who launches later with newer or more advanced capabilities. So how can we capture the value of a system (or the loss of it) as a function of its design lifetime?

In order to do so, we first **need to augment our understanding of system architect-ure(-ing).** System architecture is defined as the fundamental and unifying structure, in terms of system elements, interfaces, and constraints, of a product or a process [Maier and Rechtin, 00]. System architecting is traditionally viewed as a matching between two (vector) quantities, resources and system performance. One traditional design paradigm fixes the amount of available resources and attempts to optimize the system performance given this constraint. The other approach constrains the system performance to a desired level and strives to find a design that will achieve this performance at minimal cost [de Weck, 01]. The first approach operates with–and attempts to maximize–a **performance per unit cost** metric; the second approach seeks to minimize a **cost per function** (or performance) metric. In order to (quantitatively) discuss issues related to the **design lifetime, a fundamental component of system architecture although we cannot see it or touch it, it is indispensable that we view in an architecture the flow of service (or utility) that the system will provide over a given period of time.** We will therefore introduce **cost, utility, and value per unit time metrics** in order to guide the selection the design lifetime. This is done in the following sections.

5.3 Satellite Cost per Operational Day and Cost Profile as a Function of the Design Lifetime

In the previous chapter, we explored the effects of the design lifetime requirement on spacecraft mass and cost to IOC. We first examined how different subsystems scale with the design lifetime, then integrated these results in order to generate spacecraft mass and cost profiles as a function of the design lifetime. A typical example of a cost to IOC profile is given in Figure 5-2.
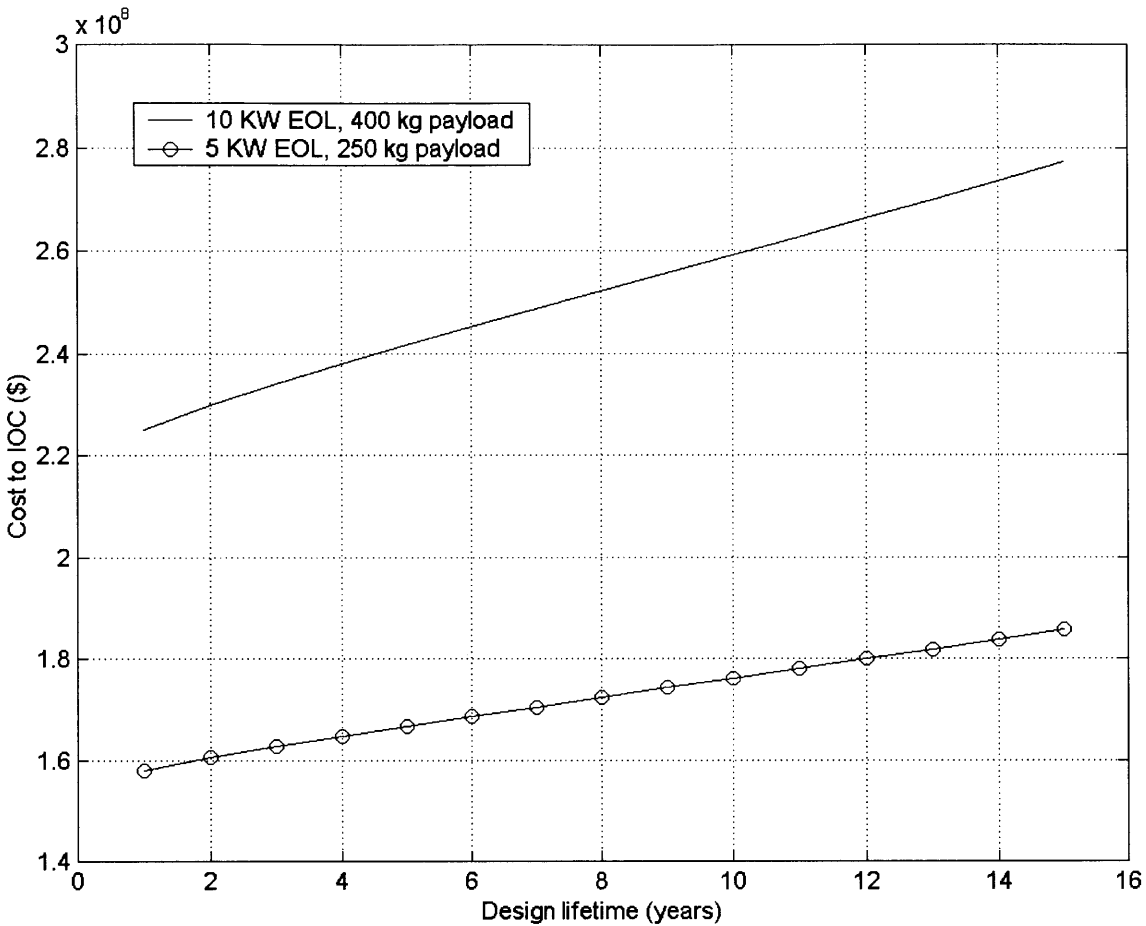
Fig. 5-2. Cost to IOC as a function of the design lifetime requirement (spacecraft in GEO, mission reliability = 95%, GaAs cells, Ni-H2 batteries).

A cost per operational day metric was defined that corresponds to uniformly amortizing the spacecraft cost to IOC over its design lifetime (without accounting for the time value of money).

$$\text{Cost}_{/day} = \frac{\text{cost to IOC}}{\text{Design lifetime (days)}} = \frac{C(T_{Life})}{T_{Life}} \tag{5-1}$$

Within the interval of the design lifetime considered, the cost per operational day decreases monotonically, as can been seen in Figure 5-3. **In the absence of other metrics**, this behavior of the cost per operational day justifies pushing the boundary of the design lifetime and designing spacecraft for increasingly longer periods. It also suggests that a customer is always better off requesting the contractor to provide the maximum design lifetime:

116

$$T_{Life} = T_{Life-\max} \qquad\qquad\qquad (5-2)$$

In the following section, we will prove that this is incorrect, and provide a way to compute spacecraft optimal design lifetime under various conditions.
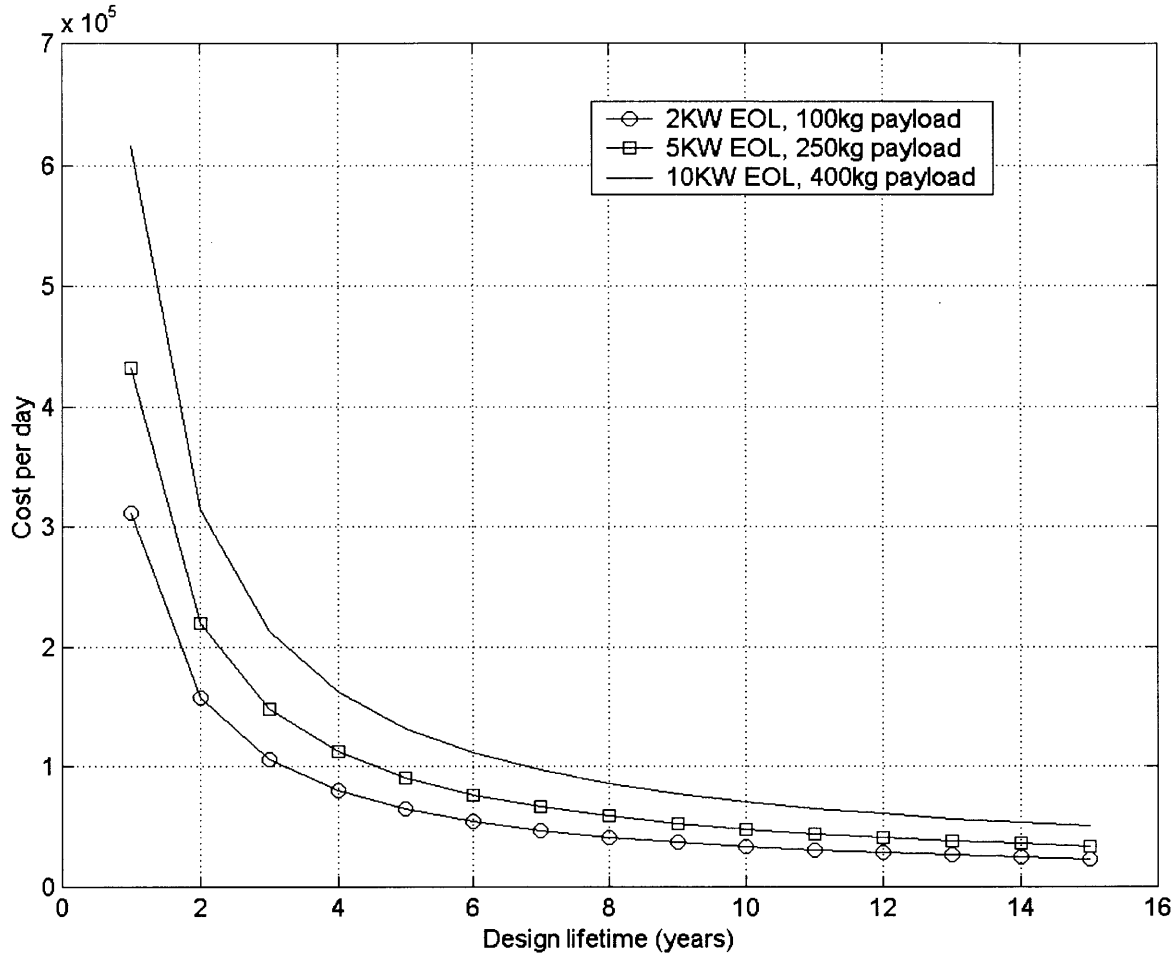
Fig. 5-3. Spacecraft cost per operational day as a function of the design lifetime (same parameters as in Fig. 5-2).

5.4 Value of a System Architecture as a Function of its Design Lifetime

In order to specify the design lifetime requirement, we need to be able to express the present value of a system as a function of its design lifetime. We propose Eq. 5-3 as a mean for capturing this value.

$$V\left(T_{Life}\right) = \int_{0}^{T_{Life}} \underbrace{[u(t) - \theta(t)]}_{\text{market analysis \& operations}} \times e^{-rt}dt - \underbrace{C\left(T_{Life}\right)}_{\text{Engineering \& cost analysis}} \qquad (5\text{-}3)$$

$\begin{cases} V(\text{T}_{Life}) : \text{Expected present value of a system architecture as a function of its design lifetime} \\\\ u(t) \quad : \text{Utility rate of the system (e.g., revenues per day for a commercial system)} \\\\ \theta(t) \quad : \text{Cost of operating the system per day} \\\\ C(\text{T}_{Life}) : \text{System cost profile as a function of its design lifetime} \\\\ r \quad\quad : \text{Discount rate} \end{cases}$

Eq. 5-3 is analogous to the continuity equation (or conservation of mass) in fluid dynamics:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho U) = 0 \qquad (5\text{-}4a)$$

in its local form, or in its integral form as follows:

$$\frac{\partial}{\partial t} \oiiint_V \rho \, dV + \oiint_S \rho \, U dS \qquad (5\text{-}4b)$$

$\begin{cases} \rho \quad : \text{Fluid density} \\\\ V \quad : \text{Control volume} \\\\ U \quad : \text{Flow velocity vector} \\\\ dS \quad : \text{Elemental surface area vector} \\\\ \nabla \cdot \quad : \text{Divergence of a vector field} \end{cases}$

The analogy between the two equations is graphically illustrated in Figure 5-4. The **control volume** becomes **a time bin** (the design lifetime), the flow entering the control volume is analogous to the revenues generated during the time considered, and the flow exiting the volume

118

corresponds to the cost of designing the system for this period of time plus the cost to operate it for the same period.

	Flow in	Accumulation	Flow out
Fluid dynamics	$\iint_{S_1} \rho\, U_1 dS_1$	$\frac{\partial}{\partial t} \oiiint_V \rho\, dV$	$\iint_{S_2} \rho\, U_2 dS_2$
System architecture	$\int_0^{T_{Life}} u(t) e^{-rt} dt$	$V(T_{Life})$	$C(T_{Life}) + \int_0^{T_{Life}} \theta(t) e^{-rt} dt$
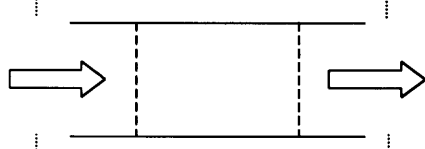
Fig. 5-4. Analogy between the expected present value of a system architecture as a function of its design lifetime (Eq. 5-3) and the continuity equation in fluid dynamics.

Two time characteristics can be readily derived from Eq. 5-3: the minimum design lifetime for a system to become profitable, and the time of operations for a system to break even given a design lifetime. These are detailed below.

5.4.1 Minimum design lifetime for the system to become profitable

The minimum design lifetime for a system to become profitable can be readily computed by setting the expected present value, $V(T_{Life})$, equal to zero:

$$V\left(T_{Life-min}\right)= \int_{0}^{T_{Life-min}}[u(t)-\theta(t)]\times e^{-rt}dt - C\left(T_{Life-min}\right)= 0$$

$$V\left(T_{Life}\right)> 0 \quad \text{for} \quad T_{Life} > T_{Life-min}$$

(5-5)

While technical considerations limit the upper bound of system design lifetime, the lower bound on the design lifetime is dictated by economic (value) considerations, and is given by the solution to Eq. 5-5. The dynamics of $T_{Life-min}$ and the parameters driving it will be discussed shortly. It should be noted that the minimum design lifetime for a system to become profitable is **NOT** identical to the "time to break even". This second time characteristic of a system is discussed below.

5.4.2 Time to break even given a design lifetime

The time for a system to break even is given by the solution of Eq. 5-6 in which T_{Life} is fixed. In other words, once the system's design lifetime is specified, time is allowed to vary until the discounted revenues cover the cost to design the system for T_{Life}, $C(T_{Life})$, in addition to the discounted cost to operate the system until $T_{break-even}$.

$$V\left(T_{Life}, T_{break-even}\right)= \int_{0}^{T_{break-even}}[u(t)-\theta(t)]\times e^{-rt}dt - C\left(T_{Life}\right)= 0 \qquad (5-6)$$

It is only when the system design lifetime is equal to $T_{Life-min}$, i.e., when the system pays off its expenses at the end of its mission, that the time to break even and the minimum design lifetime are identical.

$$T_{break-even} = T_{Life-min} \quad \text{when} \quad T_{Life} = T_{Life-min} \qquad (5-7)$$

No for-profit company would want to acquire or field such systems (zero-profit system). As the design lifetime T_{Life} increases above $T_{Life\text{-}min}$, the cost to design such a system increases:

$$C\left(T_{Life}\right) > C\left(T_{Life-min}\right) \quad \text{for} \quad T_{Life} > T_{Life-min} \tag{5 - 8}$$

Consequently, the time to break even increases as more revenues are required to cover the additional cost $\Delta C(T_{Life}) - \Delta C(T_{Life\text{-}min})$, as can be seen from Eq. 5-6. The moral of the story is:

$$T_{break-even} > T_{Life-min} \quad \text{when} \quad T_{Life} > T_{Life-min} \tag{5 - 9}$$

The comparison between the time to break even and the minimum design lifetime is summarized in Table 5-1.

Table 5-1. Time to break even, $T_{break\text{-}even}$, and minimum design lifetime, $T_{Life\text{-}min}$.

When	$T_{Life} < T_{Life\text{-}min}$	$T_{Life} = T_{Life\text{-}min}$	$T_{Life} > T_{Life\text{-}min}$
	$T_{break\text{-}even}$ does not exist	$T_{break\text{-}even} = T_{Life\text{-}min}$	$T_{break\text{-}even} > T_{Life\text{-}min}$

It is fair at this point to ask the following question: Assume the management of a company that wishes to acquire a large complex system wants to break even in $T_{break\text{-}even}$ years, what is the constant revenue per day u_0 in order to do so? The answer is readily given by Eq. 5-10:

$$\int_0^{T_{break-even}} u_0 \times e^{-rt} dt = C\left(T_{Life}\right) + \int_0^{T_{break-even}} \theta(t) \times e^{-rt} dt \tag{5 - 10}$$

Therefore

$$u_0 = r \times \frac{C\left(T_{Life}\right) + \int_0^{T_{break-even}} \theta(t) \times e^{-rt} dt}{1 - e^{-rT_{break-even}}} \tag{5 - 11}$$

Assuming the cost to design the system is larger than the cost to operate it, i.e., $C(T_{Life}) \gg \int_0^{T_{break-even}} \theta(t) \times e^{-rt} dt$ and recalling that $e^x = 1 + x + \varepsilon(x^2)$, we get:

$$u_0 \approx \underbrace{\left[\frac{C(T_{Life})}{T_{Life}}\right]}_{\substack{cost\ per \\ operational\ day}} \times \left(\frac{T_{Life}}{T_{break-even}}\right) \qquad (5\text{-}12)$$

This result illustrates the importance of the cost per operational day metric, and can prove useful in feasibility studies or back-of-the-envelope calculations. For instance, assume a company that is acquiring a \$100M system designed for ten years wishes to amortize its investment in two years. In order to do so, the company should guarantee revenues per day at least five times more than the system's cost per operational day:

$$u_0 \approx \left(\frac{100 \times 10^6}{10 \times 365}\right) \times \frac{10}{5} \approx \$55,000 / day$$

Conversely, if market analysis indicates that the service provided by this system can at best generate \$30,000/day, considering the market size and the presence of other players in this market, then the time to amortize the investment is:

$$T_{break-even} \approx \left(\frac{100 \times 10^6}{10 \times 365}\right) \times \frac{10}{30,000} \approx 9.1 years$$

It is likely, given this result, that the senior management of the company will reconsider acquiring the system with its ten years design lifetime.

5.5 Optimal Design Lifetime: Matching a System's Lifetime with the Dynamical Characteristics of its Environment

Is there an optimal design lifetime for a system architecture that maximizes its value, as given in Eq. 5-3? If so, how is it related to the dynamical characteristics of the environment in which the system is operating, such as the volatility of the market the system is serving in the case of a commercial venture, or the evolution of the system's technology base? This section addresses these questions.

It is intuitively sound for a designer to consider **matching** a system's **design lifetime** with the **dynamical characteristics of the environment** the system will operate in. It is inappropriate for example to field a system designed for twenty years when the average lifecycles of the technologies embedded in the system are of the order of three years, and the system is serving a highly volatility market with a potential to dramatically change say in a couple of years (unless one fields a "flexible" design – see Chapter 3). But how can we formally capture these issues and test their correctness? Figure 5-5 presents one aspect of the system architecting process that could help clarify this matter.
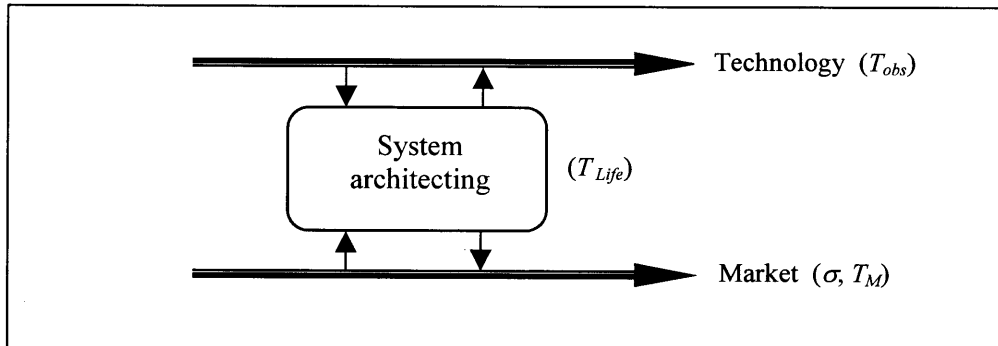
Fig. 5-5. Schematic of one aspect of the system architecting process. Adapted from Allen (1997).

As shown above, the process mediates between two streams of activities: "One of these [activities] is the development of technological knowledge, or as we more commonly call it, "technology". The other is a developing set of market needs. The basic process is then the matching of information drawn from the two streams: one stream provides market needs; the other provides technological capabilities or potential solutions to meet market needs. Both knowledge of the technology and the market are required" [Allen, 97].

Although simple, Figure 5-5 brings forth the different time scales, or clockspeeds[1], involved in the design process, namely the time constants associated with the rates of change of the product's technology base (T_{tech}), its environment or market volatility (σ, T_M), and its design lifetime (T_{Life}). The discussion in this section is articulated around these issues. For a more elaborate discussion of the system architecting process, the reader is referred to Ulrich and Eppinger (1995) on innovation in product design and development, Henderson and Clark (1990) on architectural innovation versus innovation that affects components and core concepts, and Maier and Rechtin (2000) for a holistic discussion of system architecting in general.

In the following, we first investigate the existence of an optimal design lifetime independently of technology lifecycle considerations or market volatility. We then refine our analysis by incorporating technology obsolescence issues and exploring how the system's optimal design lifetime is affected. Finally, we include market volatility in the analysis and investigate its effect on the optimal design lifetime.

5.5.1 Optimal Design Lifetime: Mathematical Formulation

Is there an optimal design lifetime for a complex engineering system? Using the notation introduced in 5-4, this question can be mathematically formulated as follows:

$$V\left(T_{Life}\right) = \int_{0}^{T_{Life}} [u(t) - \theta(t)] \times e^{-rt} dt - C\left(T_{Life}\right)$$

(5 - 13)

Is there a T_{Life}^{*} such that $V\left(T_{Life}^{*}\right) > V\left(T_{Life}\right)$ for all $T_{Life} \neq T_{Life}^{*}$?

---

[1] Fine (1998) coined the term "clockspeed" to characterize the rate of change of an industry. "I began to look at [different] industries, seeking to understand their various rates of evolution. I came to think of these rates as industry *clockspeeds*. Each industry evolved at a different rate, depending in some way on its product clockspeed [and] process clockspeed ..."

5.5.2 The Simple Case: Constant Revenues per Day, No Technology Obsolescence or Market Volatility Effects

This section explores the implications of Eq. 5-3 relating the expected present value of a spacecraft to its design lifetime, given the following **assumptions**:

Revenues per day, $u(t)$: We consider the revenues per day generated by the system constant over its design lifetime.

Spacecraft Cost profile, $C(T_{Life})$: Unless otherwise specified, we have used the cost profile of a satellite in the \$100m range with an average increase in cost to IOC of 4% per year.

Operations cost, $\theta(t)$: Mission operations are discussed in detail in Boden and Larsen (1995). They typically cost (per year) 5–15% of the spacecraft Theoretical First Unit Cost (see Chapter 4 for a discussion of various spacecraft costs). In the following, we consider constant the cost of operations per year and equal to 10% of the spacecraft cost to IOC. This can be amended in the future to include a cost profile for operations as a function of the mission phase (e.g., operations during the launch and deployment phase may require more personnel, hence be more expensive than operations after the spacecraft has been delivered to orbit and tested fully functional). This however has little effect on our results, and bear no consequences on our conceptual findings.

Discount rate, r: A discount rate of 10% is used throughout the analysis.

Figure 5-6 presents a family of curves solutions to Eq. 5-3 under the above assumptions[2]. Several observations can be made based on this plot.

1. An optimal design lifetime **exists** that maximizes the expected present value of a spacecraft as a function of its design lifetime $V(T_{Life})$. In other words, even if it is technically feasible to design a spacecraft for an extended lifetime, it is not necessarily in the best interest of the customer to ask the contractor to provide a spacecraft designed for the maximum achievable lifetime. This result, i.e., the existence of an optimal design lifetime disproves the implications of Eq. 5-2 that the customer is always better off requesting the contractor to provide a spacecraft designed for the maximum achievable

---

[2] Generating this plot the first time was a pleasant surprise. I couldn't help but find it beautiful!

lifetime. Recall that this latter conclusion was reached by considering only the monotonic decrease of the cost per operational day metric as a function of the design lifetime (see Figure 5-3).

Why do we observe this effect (existence of an optimal design lifetime)? Simply because of the time value of money: While a substantial part of the expenses, $C(T_{Life})$, are paid prior to launch, the revenues are generated at a later period, thus when discounted, are worth less than an equal amount of money spent on designing the satellite (one dollar spent today is worth more than a dollar generated in a year). Recall to this effect for instance that a constant cash flow over an infinite period of time generates a finite net present value (a dollar generated in a hundred years is worth very little today):

$$NPV = \int_0^\infty u_0 e^{-rt} dt = \frac{u_0}{r} \quad \Rightarrow \quad \text{finite!}$$
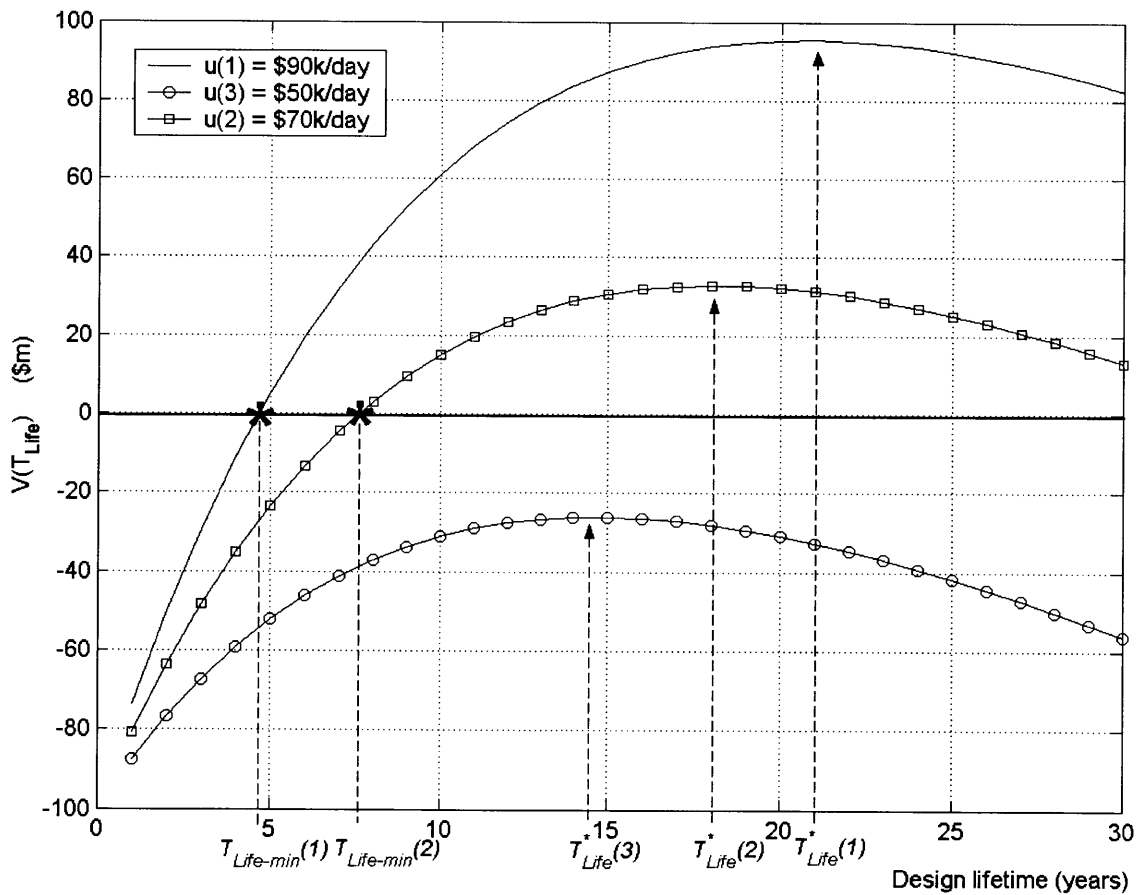
Fig. 5-6. Expected present value of a satellite as a function of its design lifetime (solutions to Eq. 5-3), assuming constant revenues per day over its design lifetime.

126

2. The optimal design lifetime T_{Life}^* increases as the expected revenues per day increase (from 14 to 21 years as the revenues increase from \$50k/day to \$90k/day). In other words, the more a customer expects to generate revenues from a system, the longer he or she would want the system to remain operational. This of course is an intuitive result; Eq. 5-13 and Figure 5-6 provide a quantitative basis for it.

3. A minimum design lifetime as defined above, $T_{Life-min}$, exists in some cases and decreases as the expected revenues per day increase. This again is intuitive and reflects the fact as the expected revenues per day increase, less time is required to amortize an initial investment. The minimum revenues per day u_{0-min}, required in order for a $T_{Life-min}$ to exist is given by Eq. 5-13 in which T_{Life}^* is equal $T_{Life-min}$ and $V(T_{Life}^*)$ is set equal to zero. In other words, the maximum expected present value of the system is set equal to zero and the solution to Eq. 5-3 is tangent to $V(T_{Life}) = 0$. In the given case, we have:

$$u_{0-min} = \$59,000/day \quad \text{and} \quad T_{Life-min} = T_{Life}^* = 16 \text{ years}$$

4. An optimal design lifetime can exist but for which the system is not profitable, as in the case above for $u_0(3) = \$50k/day$. This in fact will be the situation whenever $u_0 < u_{0-min}$. In other words, even if a system is fielded with the knowledge that it will not be profitable, it can still be designed for a period of time such that the losses are minimized (as opposed to maximizing its profits).

Table 5-2 summarizes the results discussed in this section and their implications.

Table 5-2. Summary of results and implications of Eq. 5-3 in the case of constant revenues per day over the system's design lifetime.

Results	Implications
An optimal design lifetime exists.	Even if it is technically feasible to design a spacecraft for a longer lifetime, it is not necessarily in the best interest of a customer to do so.
The optimal design lifetime increases as the expected revenues per day increase.	The more a customer expects to generate revenues from a system, the longer he or she would want the system to remain operational.
A minimum design lifetime exists for the system to become profitable, and it decreases as the expected revenues per day decrease.	A minimum revenue per day must be guaranteed for the system to be profitable. In order to decrease the minimum design lifetime for the system to be profitable, and consequently the time to break even, more revenues per day must be sought.
An optimal design lifetime can exist for which the system is not profitable.	Even if a system is fielded with the knowledge that it will not be profitable, it still can be designed for a period of time such that the losses are minimized (as opposed to maximizing its profits).

Sensitivity Analysis

We now perturb the assumptions underlying the analysis above and explore the impacts on our findings. Three parameters affect the solution of Eq. 5-3, namely the cost of operations $\theta(t)$, the discount rate r, and the satellite cost profile $C(T_{Life})$.

Sensitivity to cost of operation: We have previously considered the cost of operations per year to be constant and equal to 10% of the spacecraft cost to IOC. As can be seen from Eq. 5-3, $\theta(t)$ acts upon the optimal design lifetime in the opposite way to the revenues per day $u(t)$. As we increase the cost of operation per year from 5% to 15% of the spacecraft cost to IOC (a reasonable range that covers most cases), the optimal design lifetime decreases by less than one year. This change is minor and inconsequential to the decision of selecting spacecraft design lifetime.

Sensitivity to discount rate: We have previously assumed a discount rate of 10%, a value typically used in the aerospace industry [Wertz and Larson, 99]. As we vary the discount rate, keeping all other parameters constant, we note a change in the location of the optimal design lifetime. The optimal design lifetime decreases as the discount rate increases, as shown in Table 5-3.

Table 5-3. Effect of the discount rate on optimal design lifetime ($u_0 = \$75k/day$).

Discount rate	Optimal design lifetime
8%	22 years
10%	19 years
12%	16 years

How can we explain this effect? Again by the time value of money: The present value of a dollar generated in a year decreases as the discount rate increases ($0.93 when $r = 8\%$, and $0.89 when $r = 12\%$). Increasing the discount rate therefore has the same effect on the optimal design lifetime as decreasing the expected revenues per day. Consequently, as the discount rate increases, the optimal design lifetime decreases. The impact of the discount rate on the optimal design lifetime is not negligible, as can be seen in Table 5-3. Therefore, particular attention should be paid to the discount rate when performing such an analysis.

Sensitivity to satellite cost profile: We have previously assumed a satellite cost profile in the $100M range with an average increase in its cost to IOC of 4% per year. We find that the optimal design lifetime is most sensitive to the average increase per year of the system cost profile. In other words, the more it costs to design a system for an extended period of time, the less it is profitable to do so. Figure 5-7 illustrates this result.
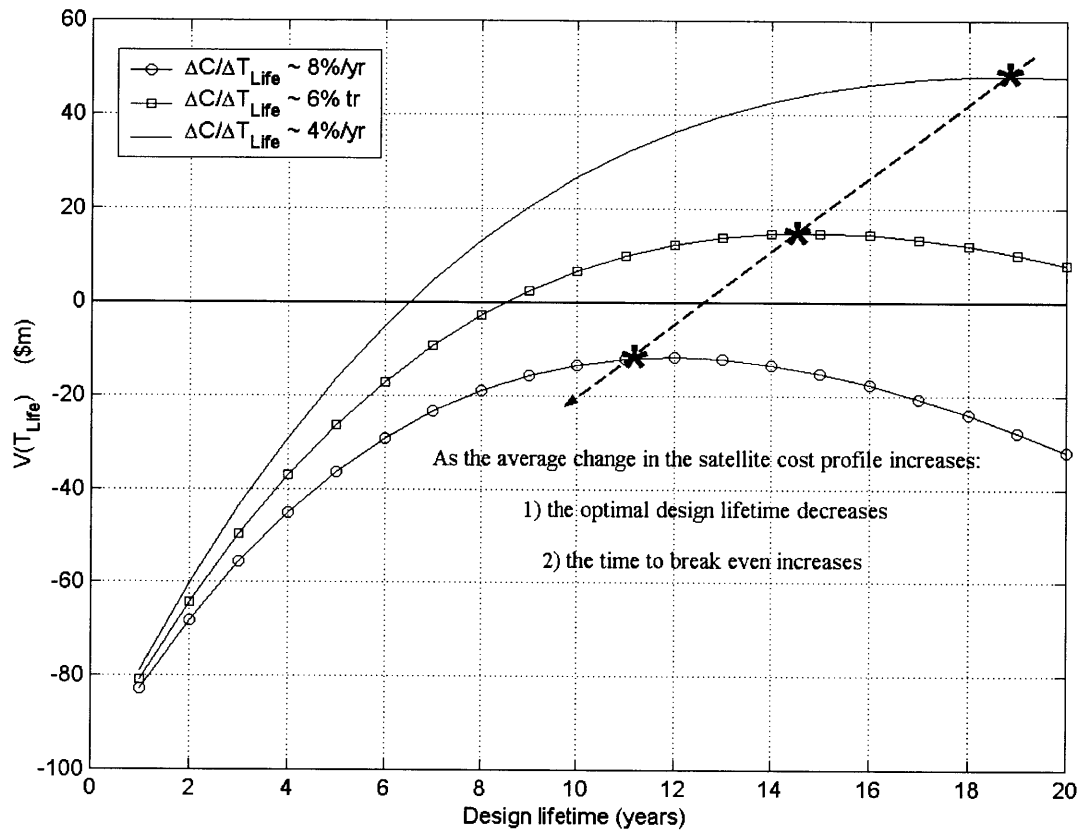
Fig. 5-7. Spacecraft optimal design lifetime as a function of the average increase in the spacecraft cost profile $(u_0 = \$75k/day)$.

5.5.3 Technology Obsolescence and Optimal Design Lifetime

In this section, we seek to answer two questions: First, how are the revenues per day generated by a system affected by the time scale associated with its obsolescence? Second, how is the optimal design lifetime affected by the system's time scale to obsolescence?

Technology obsolescence is discussed in detail in Chapter 3, The Case for Flexibility in System Design. Below is a summary of the main points necessary for our analysis.

Many complex systems, particularly aircraft and spacecraft as well as major military systems, often have design lifetimes far in excess of several life spans of many of their supporting technologies. Indeed the commercial new product cycle for many technologies, including the rapidly evolving electronic component industry, is two years or less. With systems that are designed to remain in operation for 15 years or more, obsolescence problems become critical [Hitt and Schmidt, 98].

Although **increased reliability** as well as **budgetary constraints** have **lengthened** complex **systems design lifetimes**, rapid advances in technology have dramatically shortened component life cycle. This decrease is particularly acute for electronic components but affects non-electronic components as well. The disparity between the long lifetime of systems and the short life span of components embedded in the system requires careful consideration of issues of obsolescence and upgradability, or the ability to modify a system after it has been fielded in a timely and cost-effective way.

Figure 5-8 shows a typical component's progress towards obsolescence and introduces the concept of time to obsolescence.
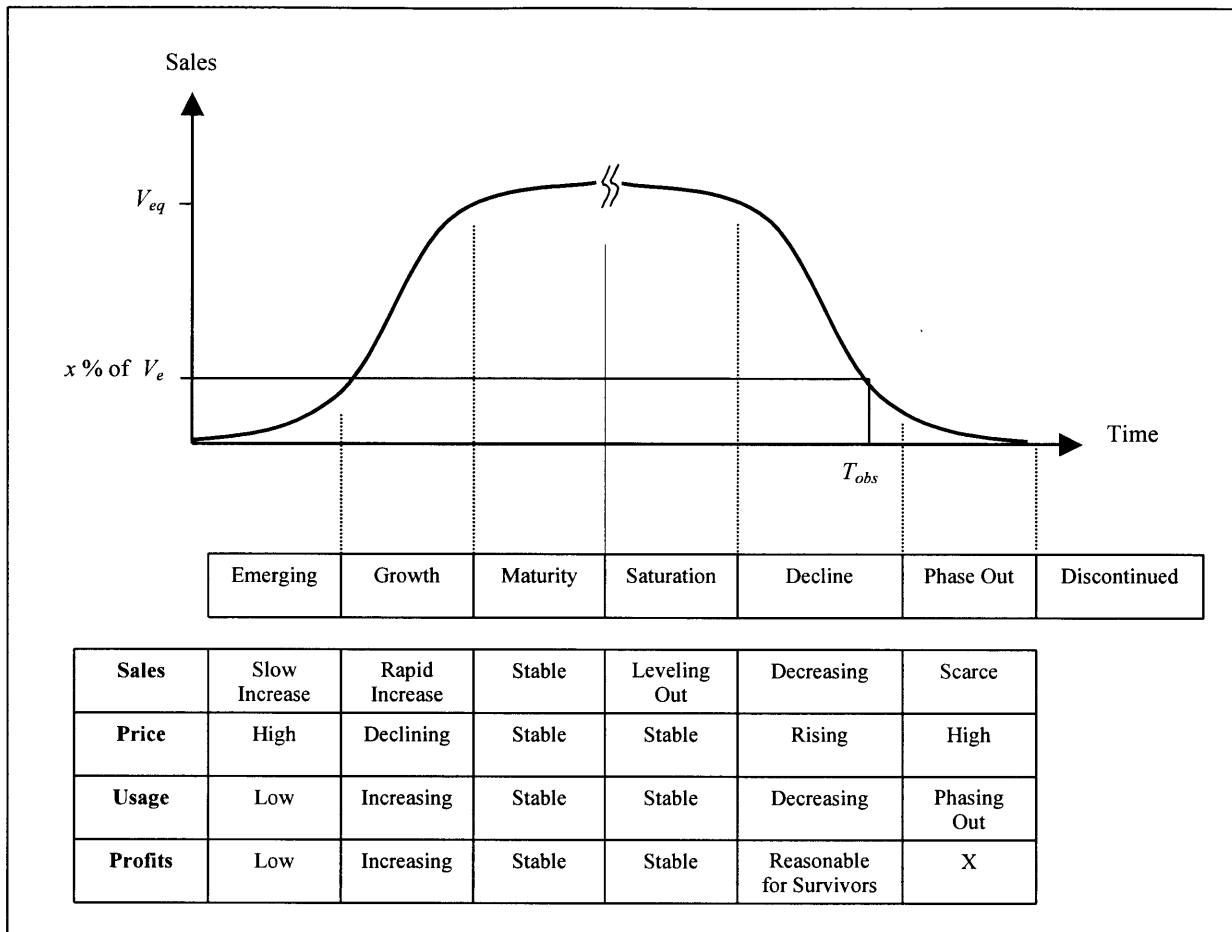
Fig. 5-8. Typical component life cycle and phases. Adapted from the ANSI/EIA-724-97 Product Life Cycle Data Model [Hatch, 00]. Time to obsolescence T_{obs} can be defined in relation to sales volume.

One way of defining Time to Obsolescence of a component, T_{obs}, is in relation to sales volumes, as illustrated above. The challenge, however, is in relating component obsolescence to system's (or mission) obsolescence. This topic is worthy of its own thesis. In this section, we assume that obsolescence prediction can be performed at the system's level, and that the revenues per day generated by a system are affected by its time to obsolescence according to Eq. 5-14:

$$u(t) = u_0 \times \exp\left[-\left(\frac{t}{T_{obs}}\right)^2\right]$$

(5 - 14)

We will first treat T_{obs} of a system as a deterministic variable, then as a random variable with an associated distribution function $P(T_{obs})$. The main difference between our (system) model of

132

obsolescence (Eq. 5-14) and the ANSI model of component lifecycle (Figure 5-8) is the absence of emerging and growth phases. We justify this difference by the fact that while components are mass-produced, the systems we are considering in our analysis are dedicated and often one-of-a-kind. In addition, they reach full operational capability in a time much shorter than their design lifetime, thus the emerging and growth phases are very short compared with their design lifetime. Our model can be easily amended to include a short initial period of rise in the revenues per day. This however has little effect on the results.

Time to Obsolescence: The Deterministic Case

Figure 5-9 illustrates Eq. 5-14 for various estimates of system's Time to Obsolescence. It shows, as expected, that the revenues per day decrease faster as the Time to Obsolescence decreases. This answers the first question we set to investigate in this section, namely, the relationship between the revenues per day and a time scale associated with the system's obsolescence.
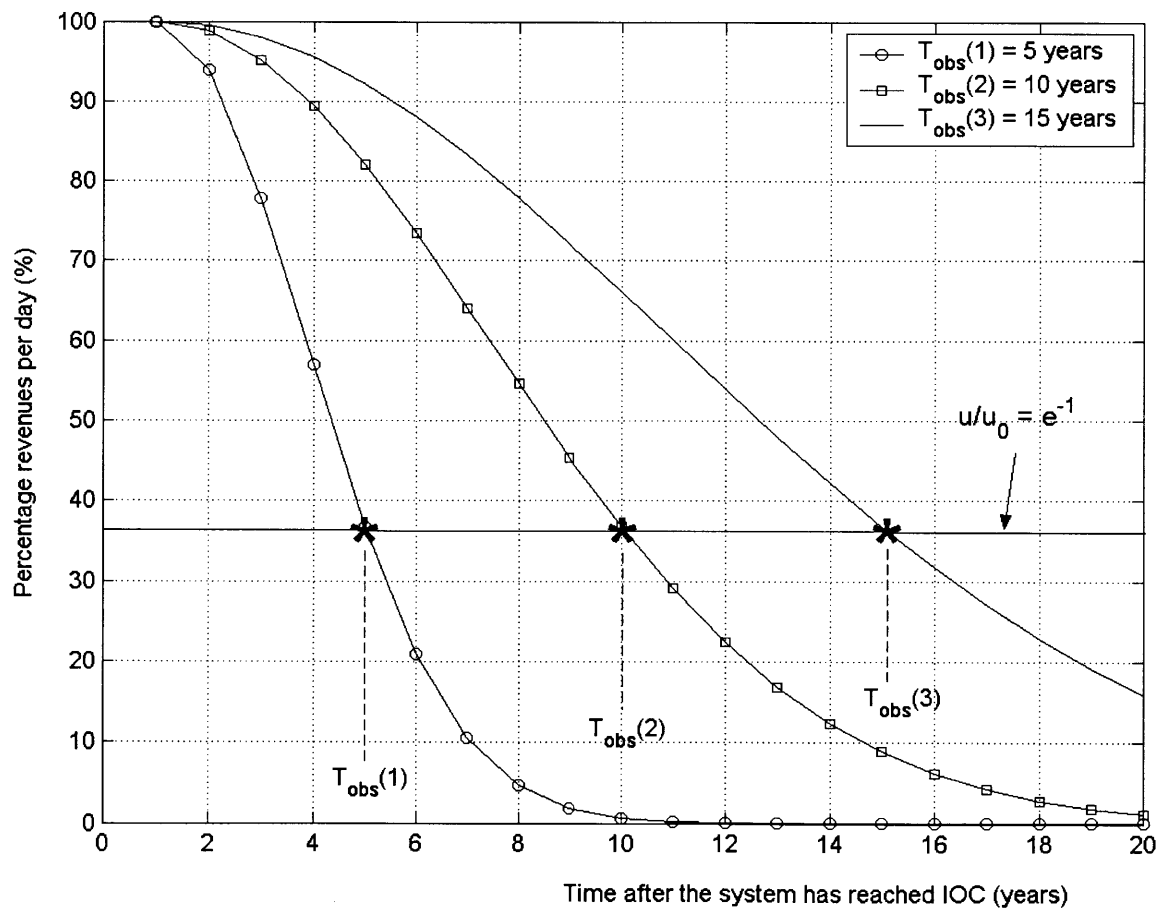
Fig. 5-9. Percentage revenues per day as a function time for various estimated system's Time to Obsolescence.

133

Given our model of evolution of the revenues per day as a function of the system's obsolescence, how is the system's optimal design lifetime affected?

We can readily answer this question by replacing the revenues per day in Eq. 5-3 by our model in Eq. 5-14. The same assumptions as above are considered regarding the discount rate and the cost of operations. Figure 5-10 presents the expected present value of a satellite as a function of its design lifetime assuming the revenues generated per day are affected by technology obsolescence and are given by Eq. 5-14.
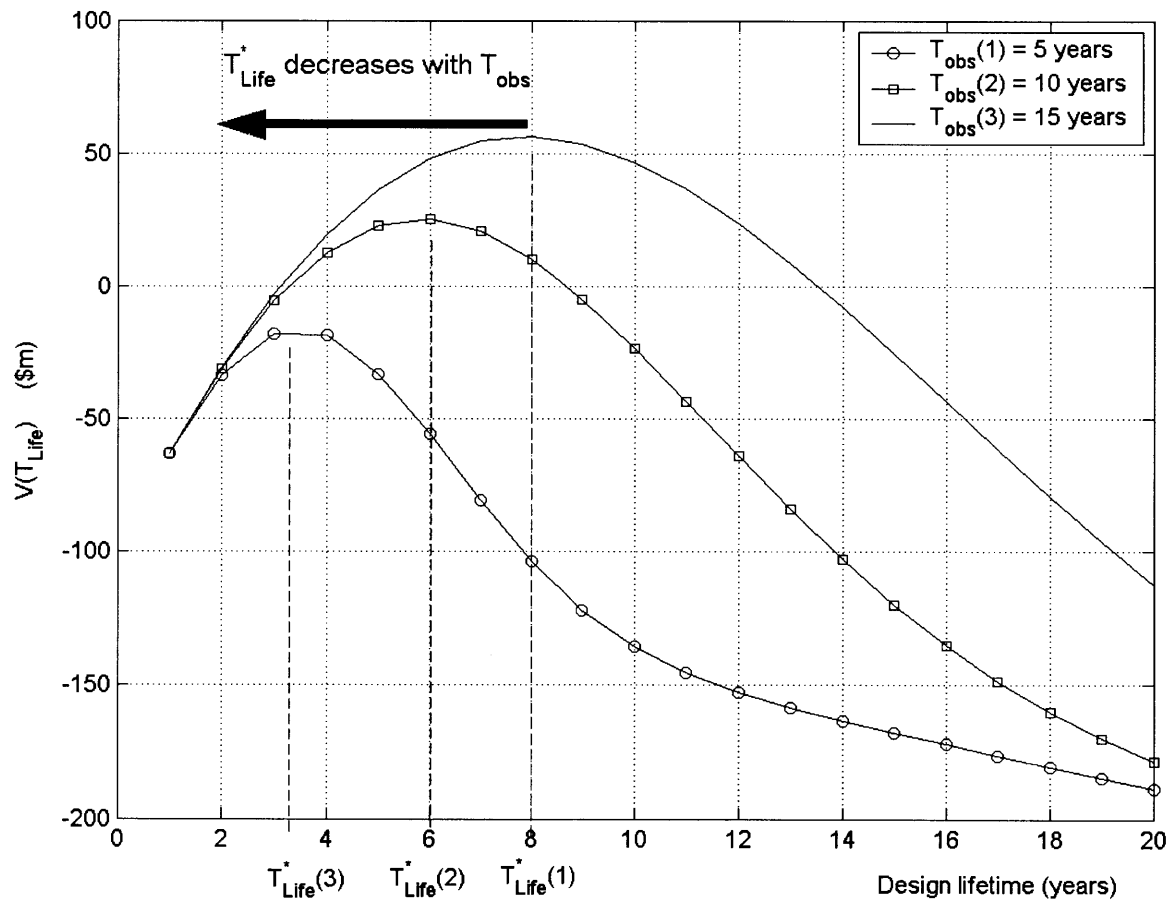
Fig. 5-10. Expected present value, $V(T_{Life})$, of a satellite as a function of its design lifetime (solutions of Eq. 5-3), assuming revenues per day affected by system's obsolescence (Eq. 5-14). Optimal design lifetime deceases as the expected Time to Obsolescence decrease (u_0 = $120k/day, $\Delta C/\Delta T \sim 4\%$/year).

We note that the optimal design lifetime deceases (from 8 to 3.5 years) as the expected system's Time to Obsolescence decreases (from 15 to 5 years). In other words, the sooner a customer expects a system to become obsolete, the less time they would want it to be designed for. While this result is intuitive, Eq. 5-13 and 5-14, and Figure 5-10 provide a quantitative justification for it. In addition, we note a significant decrease in the optimal design lifetime between the constant revenues assumption (no technology obsolescence) and the case where obsolescence considerations are factored into the analysis (from $T_{Life}^* > 20$ years to $T_{Life}^* \sim 8$ years with an expected Time to Obsolescence of 15 years). This on the one hand illustrates the importance of performing obsolescence prediction analysis, and on the other hand of **matching** a system's **design lifetime** with the **dynamical characteristics of the environment** the system will operate in, in this case the evolution of the system's technology base.

Time to Obsolescence: The Probabilistic Case

It is likely that the time to obsolescence is a random variable that can only be known through its probability density function. In Chapter 3, we modeled the Time to Obsolescence of a component as a random variable with a lognormal probability density function. The main points for this representation are recalled below. The reader is referred to Chapter 3 for more details. The lognormal distribution is applicable to random variables that are constrained by a lower bound (e.g., the Time to Obsolescence cannot be negative), and can have few large values. A third parameter τ, called the waiting time or shift parameter, defines the lower bound of the random variable. The probability density function of the expected Time to Obsolescence, $\hat{T}_{obs}$, is given in Eq. 5-15.

$$p(\hat{T}_{obs}) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)\left(\hat{T}_{obs} - \tau\right)} \exp\left[-\left(\frac{\log\left(\left(\hat{T}_{obs} - \tau\right)/m\right)}{\sigma\sqrt{2}}\right)^2\right]$$

(5 - 15)

In some cases, it is easier to obtain–or understand–the cumulative distribution function of a random variable than its probability density function. Figure 5-11 presents both a typical probability density function as well as a cumulative distribution function of the Time to Obsolescence for a microprocessor (m=1.5 years, σ= 0.8 years, τ= 0.5 years).
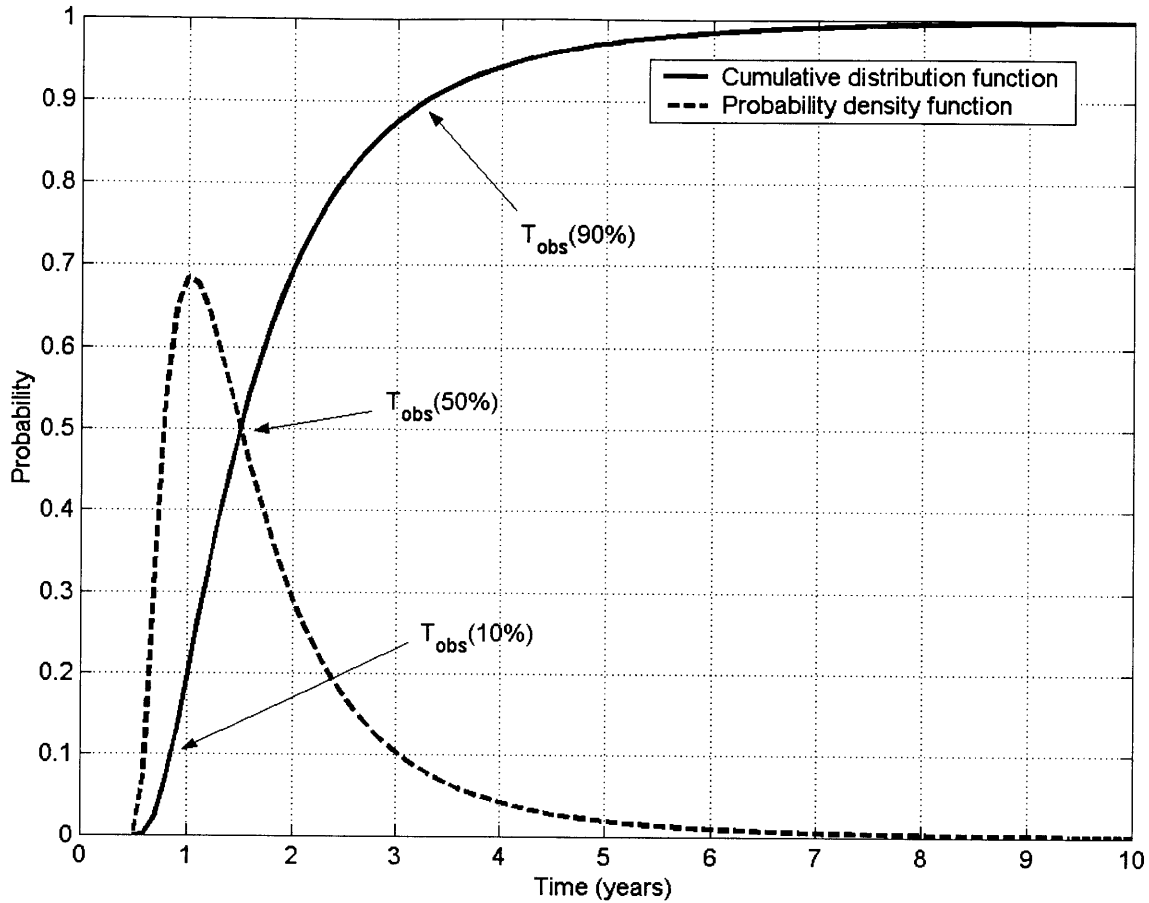
Fig. 5-11. Cumulative distribution function and probability density function of the Time to obsolescence for a microprocessor (m=1.5 years, σ= 0.8 years, τ= 0.5 years).

Now, assuming a similar distribution function for the Time to Obsolescence of a system (not a component), what is the optimal design lifetime of the system? This problem is identical to the one formulated in Eq. 5-13 in which we sought to maximize the expected present value of a system architecture as a function of its design lifetime, given the profile of the expected revenues per day (Eq. 5-14) and the probability density function of the Time to Obsolescence of the system (Eq. 5-15). This question can be readily cast into the following mathematical format:

$$V^*\left(T_{Life}\right)= \max_{T_{Life}} \int_{T_{Life}} \int_0^\infty \left\{ p\left(\hat{T}_{obs}\right)\times\left[u_0 \exp\left(-\left(\frac{t}{\hat{T}_{obs}}\right)\right)^2 -\theta(t)\right]\times d\hat{T}_{obs} \right\}\times e^{-rt}dt \; - \; C\left(T_{Life}\right) \qquad (5 - 16)$$

This problem will be explored in future work. However, a fundamental point is made by considering the following simple hypothetical situation:

136

Assume that the Time to Obsolescence of the system can only take two values, 10 years or 15 years with equal probability. Under similar conditions to those in Figure 5-11, the expected present value of the system designed for 6 years (i.e., assuming a Time to Obsolescence = 10 years) is equal to $37M, while it is only worth $31M if the system is designed for 8 years (i.e., assuming a Time to Obsolescence of 15 years). These results are summarized in Table 5-4.

Table 5-4. Expected present value of a system architecture and optimal design lifetime as a function of the expected Time to Obsolescence. Based on Fig. 5-11.

T_{obs_des} (years)	T_{Life}^{*} (years)	$\hat{T}_{obs}$ (years)	V ($m)	$p(\hat{T}_{obs}=10) \times V_{10} + p(\hat{T}_{obs}=15) \times V_{15}$
10	6	10	$V_{10} = 25$	$37M
		15	$V_{15} = 49$	
15	8	10	$V_{10} = 7$	$31M
		15	$V_{15} = 55$	

This result suggests that a system architect is sometimes better off fielding a system designed for a short lifetime and considering the option of upgrading the system at a later period depending on how the situation evolves, instead of designing the system upfront for an extended lifetime and running the risk of reduced profits or even large losses if the system actually turns out to be obsolete earlier.

While it is relatively easy to repair, maintain, or upgrade systems that are physically accessible, doing the same with satellites is more challenging. On-orbit servicing provides a way of physically accessing and upgrading satellites. The value of upgrading a satellite, as seen from a customer perspective, can only be captured by performing the type of analysis introduced in this section. These issues are discussed in the following chapter.

5.5.4 Market Volatility and Optimal Design Lifetime

In the previous section, we explored the impact of technology obsolescence on the selection of an optimal design lifetime for a satellite. In order to do so, we first assumed a relationship between the system's Time to Obsolescence and the revenues generated per day by the system. We then computed its optimal design lifetime as specified in Eq. 5-13. No considerations however were given to the dynamical characteristics of the market the system is serving. It is clear though that the revenues per day (or utility rate) generated by the system are intrinsically related to the volatility of the market the system is serving (or the uncertainty characterizing the system's environment): For example, as the market for a given service increases, it is likely that the revenues generated per day from a system providing this particular service will increase. Conversely, as the market for this service decreases, the revenues generated per day by the system will decrease.

$$u(t) = u(\text{market volatility, technology obsolescence, ...}) \quad\quad (5 - 17)$$

In this section, we investigate the effect of market volatility on the selection of a spacecraft optimal design lifetime (without considerations of technology obsolescence). The key for addressing this matter is the impact of the market volatility on revenues generated per day.

In order to do so, we first need a relationship between market volatility and time. We can use to this effect a graphic representation that is widely used in the real options literature, the cone of uncertainty. This is a simple and intuitive representation of the relationship between uncertainty and time [Amram and Kulatilaka, 99]. The cone represents how an uncertain parameter may evolve in the future. The apex of the cone represents the present (e.g., the observed present value of a stock). As one looks further into the future, there is more uncertainty about the forecast. Consequently, the value of the parameter can fall within an increasingly larger interval. We thus obtain the conic shape of the "cone of uncertainty". This is captured in Figure 5-12.

138
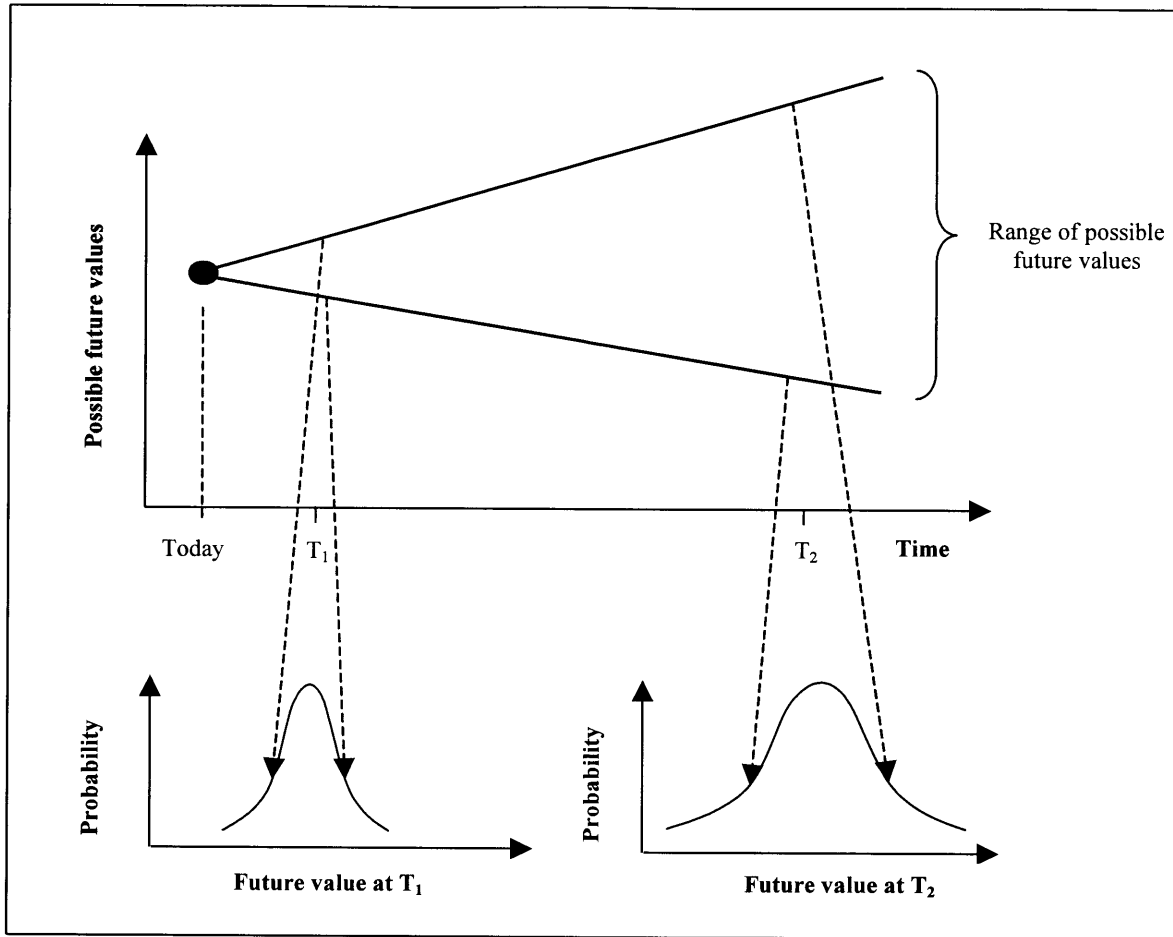
Fig. 5-12.Two related views of the propagation of uncertainty with time. The cone of uncertainty (top) represents future values of a project. Slicing through the cone, we observe (below) the probability density function at each point in time. As one looks further, there is more uncertainty about the forecast. This is represented by an increasing standard deviation of the pdf with time. Adapted from [Amram and Kulatilaka, 99].

The analysis is based on the two following assumptions:

1. First, we assume that the value of the market the system is serving has a lognormal probability density function. This is a standard result in real option theory; it results from the assumption that the future value of a real asset behaves as a financial stock, therefore its rate of change can be described as a diffusion process (random walk) with volatility σ (the standard deviation increases as $\sigma\sqrt{t}$). The reader is referred to Trigeorgis (1996) for a comprehensive discussion of the diffusion process in modeling the dynamics of the value of real and financial assets.

139

2. Second, we assume that the revenues generated per day by a system serving this market are directly correlated with the dynamics of the market, i.e., the same volatility characterizes the market and the revenues generated per day (from this market).

Figure 5-13 represents a modified cone of uncertainty (90% confidence interval) for the expected revenues per day generated by a system serving a market characterized by various volatilities. Since this analysis is performed prior to launch, the apex of the cone is not visible, and there is uncertainty about the revenues per day at T = 0 year (after the system has reached IOC).
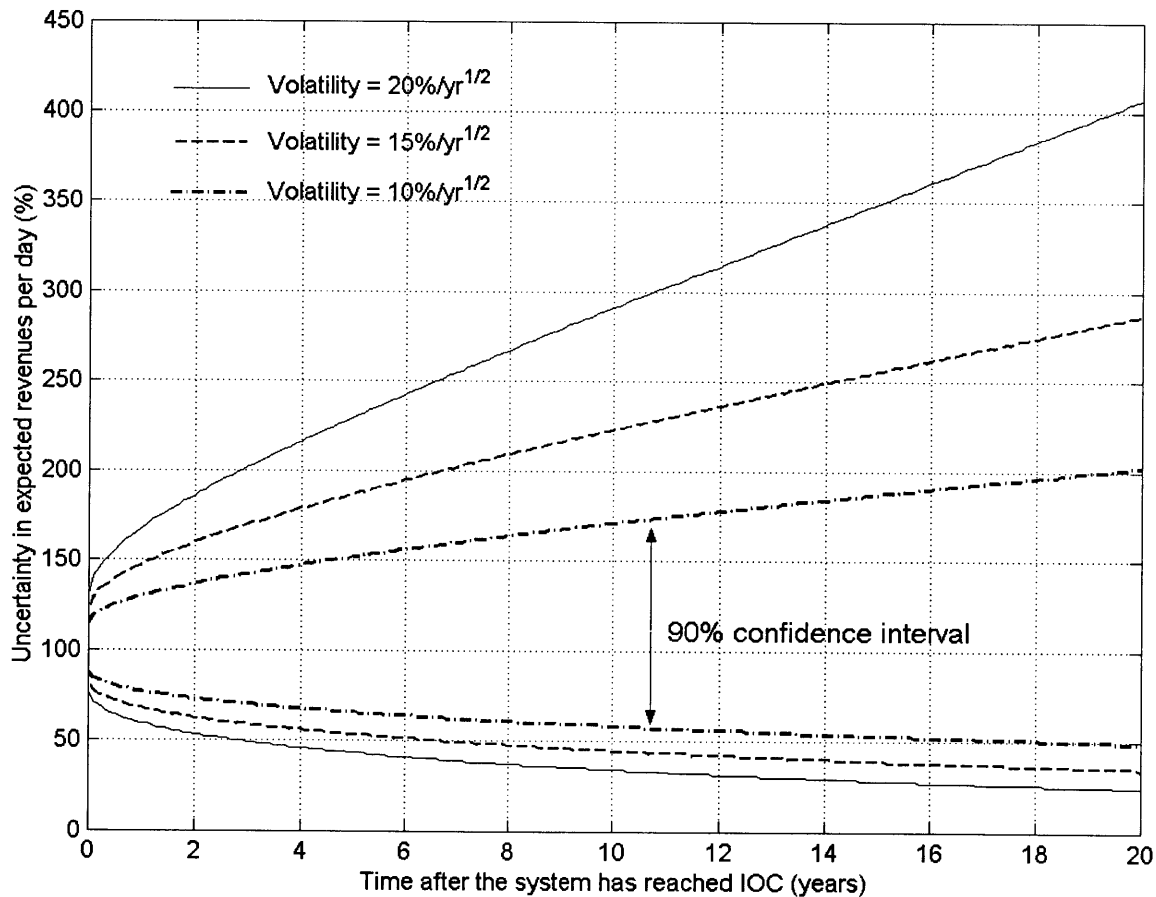
Fig. 5-13. Cone of uncertainty of the expected revenues per day for different market volatilities (90% confidence interval). It is unlikely that the revenues per day will fall outside the cone.

For a risk-averse decision-maker (one that expects worst case scenario and designs for it), we can compute the expected present value of satellite as a function of the design lifetime for various market volatilities. This is done by substituting the expected revenues per day shown in Figure 5-13 into Eq. 5-3. The results are presented in Figure 5-14.

140
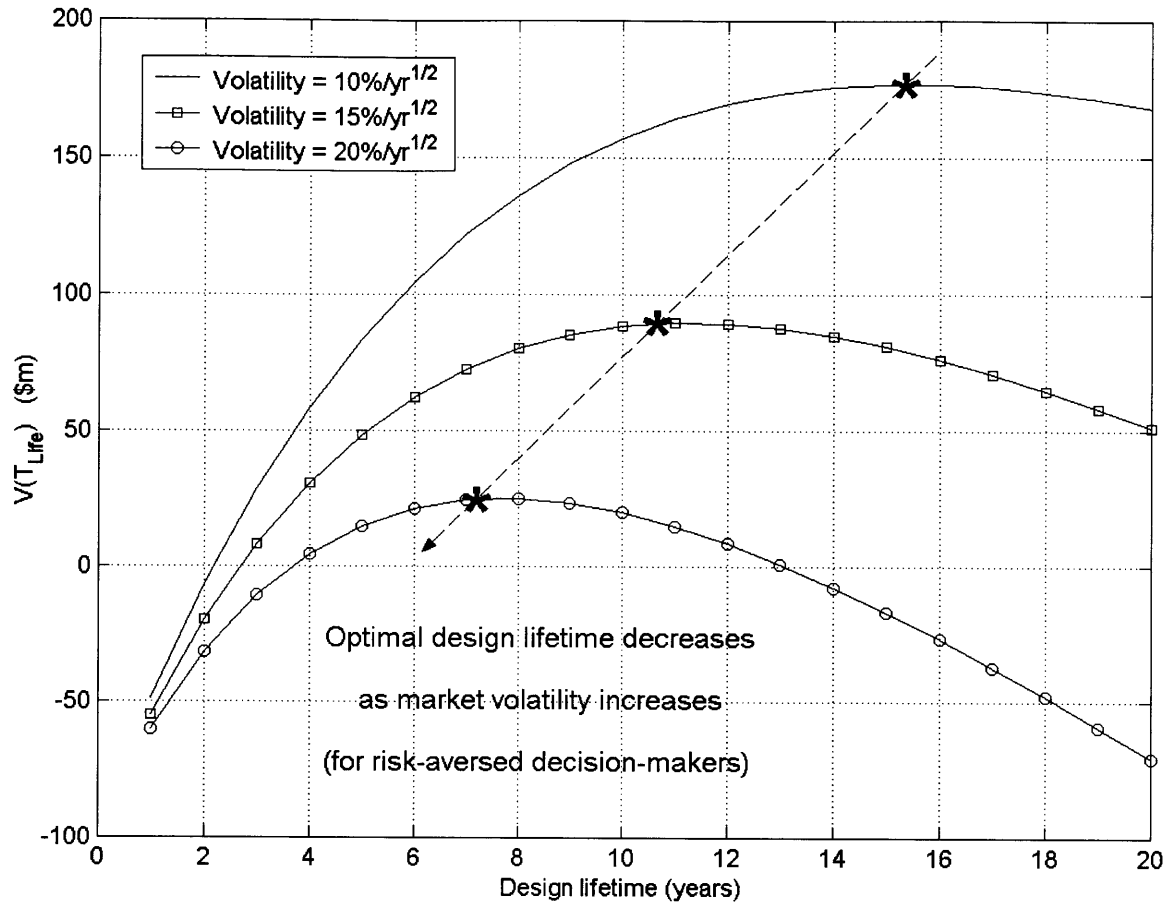
Fig. 5-14. Expected present value of a satellite as a function of its design lifetime (solutions of Eq. 5-3), assuming revenues per day affected by market volatility. Optimal design lifetime deceases as the market volatility increases (u_0 = $120k/day, $\Delta C/\Delta T \sim$ 4%/year).

We note that the optimal design lifetime decreases as the market volatility increases. However having the option to expand, upgrade, or modify the system depending on how the market actually evolves becomes increasingly valuable. On-orbit servicing provides a way for physically accessing and extending a satellite design lifetime. As in the case of upgrading a satellite, the value of satellite life extension, as seen from a customer perspective, can only be captured by performing the type of analysis introduced in this section. These issues are discussed in the following chapter.

It is worth noting that the findings regarding the satellite optimal design lifetime as a function of market volatility are in accord with a fundamental lesson from the Real Options approach: that

141

there is great value in breaking up large projects in uncertain markets [Amram and Kulatilaka, 99], or staging investments in volatile environments.

5.6 Summary and Conclusions

The design lifetime requirement, in the case of satellite systems, is "assigned rather arbitrarily" with an understanding of the technical limitations and an intuition regarding the economical impacts associated with designing for longer lifetimes [Wertz and Larson, 99]. This Chapter provided an analytical framework that quantifies the expected value of a system's architecture as a function of its design lifetime, and computes the optimal design lifetime of a spacecraft (or any system for which a cost profile is established and the revenues it can generate per day estimated).

First, we proposed to augment our understanding of system architecture by considering the **design lifetime as a fundamental component of system architecture**, although we cannot see it or touch it. This led use to view in an architecture the **flow of service** (or utility) that the system will provide over a given period of time, and to introduce **cost, utility, and value per unit time metrics**.

Second, we established a fundamental equation defining the value of a system architecture as a function of its design lifetime. This equation is analogous to the continuity equation in fluid dynamics in which the control volume becomes a time bin (the design lifetime), the flow entering the control volume is analogous to the revenues generated during the time bin considered, and the flow exiting the volume corresponds to the cost of designing the system for this time bin plus the cost to operate it during the same period. From this equation, we derived several time characteristics associated with the system, namely the minimum design lifetime for a system to become profitable, the time of operations for a system to break even given a design lifetime, and the system's optimal design lifetime.

Third, we explored the existence and dynamics of an optimal design lifetime under various conditions (constant revenues per day, technology obsolescence effects, and market volatility). Several results are worth noting. First, an optimal design lifetime exists that maximizes the value of an architecture as a function of its design lifetime. This implies that even if it is technically feasible to design a spacecraft for an extended lifetime, it is not necessarily in the best interest of the customer to ask the contractor to provide a spacecraft designed for the maximum achievable lifetime. Second, an optimal design lifetime can exist but for which the system is not profitable.

This implies that even if a system is fielded with the knowledge that it will not be profitable, it still can be designed for a period of time such that the losses are minimized (as opposed to maximizing its profits). Third, the optimal design lifetime decreases as the expected system's Time to Obsolescence decreases. Finally, the optimal design lifetime decreases as the volatility of the market the system is serving increases. Overall, these results prove that it is essential for a system architect to match a system's design lifetime with the dynamical characteristics of the environment the system is operating in—unless one embeds flexibility in the system design (as defined in Chapter 2 and argued for in Chapter 3).

Finally, we saw that as the uncertainty on the system's Time to Obsolescence increases, or as the volatility of the market the system is serving increases, it becomes increasingly valuable to have the option to upgrade the system or extend its design lifetime depending on how events unfold. On-orbit servicing provides a way for physically accessing, upgrading, and or extending a satellite design lifetime. The value of satellite life extension, as seen from a customer perspective, can only be captured by performing the type of analysis introduced in this chapter. On-orbit servicing, the flexibility it provides to space systems, and its value are discussed in the following chapter.

The Blind Men and the Elephant

A poem by John Godefrey Saxe, based on a Hindu fable.

It was six men of Indostan
To learning much inclined
Who went to see the elephant
Though all of them were blind
That each by observation
Might satisfy his mind.

The first approached the elephant
And happening to fall
Against his broad and sturdy side
At once began to bawl:
"God bless me! But the elephant
Is very like a wall!"

The second, feeling of the tusk
Cried, "Ho! What have we here
So very round and smooth and sharp?
To me 'tis mighty clear
This wonder of an elephant
Is very like a spear!"

The third approached the animal,
And happening to take
The squirming trunk within his hands,
Thus boldly up and spake:
"I see," quoth he, "the elephant
Is very like a snake."

The fourth reached out his eager hand,
And felt about the knee.
"What most wondrous beast is like
Is mighty plain," quoth he;
"'Tis clear enough the elephant
is very like a tree!"

The fifth who chanced to touch the ear,
Said: "Even the blindest man
Can tell what this resembles most;
Deny the fact who can,
This marvel of an elephant
Is very like a fan!"

The six no sooner had begun
About the beast to grope
Than, seizing on the swinging tale
That fell within his scope,
"I see," quoth he, "the elephant
Is very like a rope!"

And so these men of Indostan
Disputed loud and long,
Each in his own opinion
Exceeding stiff and strong,
Though each was partly in the right,
And all were in the wrong!

So, oft in ~~theological~~ academic wars, |
The disputants, I wean,
Rail on in utter ignorance
Of what each other mean,
And prate about an elephant
Not one of them has seen[†].

[†] Now re-read the poem by replacing "elephant" with "on-orbit servicing"!

Chapter 6

Flexibility and the Value of On-Orbit Servicing:
A New Customer-Centric Perspective

"As our case is anew, so we must think anew, and act anew. We must disenthrall ourselves"

Lincoln to Congress in 1862.

6.1 Introduction

While the majority of weapon systems take advantage of logistics and maintenance support; aircraft operational lifetime and capabilities are extended through routine maintenance and payload upgrades, satellites remain the only complex engineering systems without maintenance, repair, and upgrade infrastructure.

The absence of space logistics and infrastructure, coupled with decision-makers' desire to lower satellites cost-per-operational day, leads to the design of spacecraft for the longest operational lifetime. Over the last two decades, telecommunication satellites have seen their design lifetime on average increase from seven to fifteen years. Life extension occurred simply because it became technically feasible to design for a longer lifetime. This is the case of the AT&T's Telstar 3 communications satellites based on the Hughes HS-376 bus. The satellites have ten-year design lives, as opposed to seven-year lives for earlier satellite models. Life extension was made possible by the use of improved Nickel-Cadmium batteries and the introduction of solid-state power amplifiers in place of traveling wave tubes.

Designing for the longest technically achievable lifetime, however, hampers the rapid deployment of new technologies and capabilities since new technologies and capabilities can only be provided

145

as the satellites retire. It also increases the risk that the spacecraft becomes technically and commercially obsolete before the end of its mission (for a discussion of market dynamics, technology evolution and obsolescence, see §3.3). This is graphically illustrated in Figure 6-1.
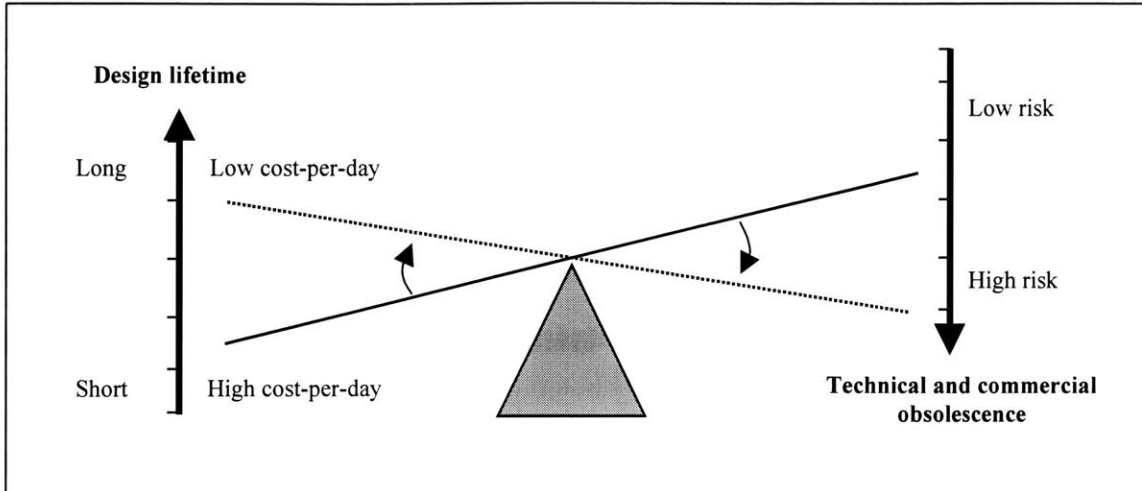
Fig. 6-1. Graphical illustration of the design lifetime trade-offs: Designing for the longest achievable lifetime decreases the satellite's cost-per-operational day. However, it increases the risk that the satellite becomes technologically and commercially obsolete before the end of its lifetime.

On-orbit servicing would provide a substantial advantage to commercial or military organizations over their competitors (or adversaries) by de-coupling the drive to lower satellites cost-per-operational day through extended design lifetime from the ability to respond quickly to changing requirements and deploying new capabilities (see Fig. 6-2). **In other words, on-orbit servicing provides flexibility to space systems.** Flexibility is defined here as the property of a system that allows it to respond to changes in its initial requirements and objectives, occurring after the system has been fielded, in a timely and cost-effective way.
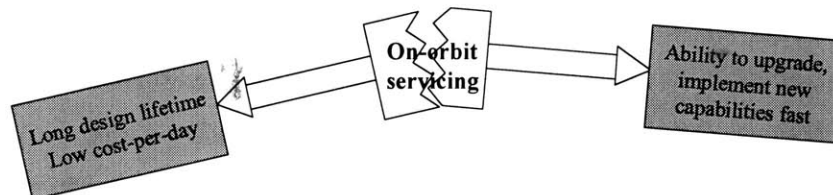
Fig. 6-2. On-orbit servicing as a solution for decoupling the drive to lower satellites cost-per-operational day through extended design lifetime from the ability to respond quickly to changing requirements and deploying new capabilities.

146

Numerous studies have been written on the subject of on-orbit servicing in the 1970's and 1980's assuming routine and economical access to space via the Space Shuttle (e.g., the Space Assembly, Maintenance, and Servicing study). Other design studies were performed more recently, establishing requirements, constraints, and technology needs of robotic on-orbit servicing, and proposing point design solutions for on-orbit servicers (SMARD, GPS servicing, etc.). Despite these efforts, fundamental questions of applicability and cost-effectiveness of on-orbit servicing remain unanswered.

This chapter proposes a new perspective on on-orbit servicing where the value of on-orbit servicing is studied independently from its cost. A framework is developed that captures the value of flexibility provided by on-orbit servicing to space systems. Several *options* are made available to space missions through on-orbit servicing (e.g., option to service for life extension, or option to upgrade) that need not be set prior to launch; they can be *exercised* after the spacecraft has been deployed, depending on how events unfold (market changes, new military contingency, etc.). It is argued that only by accounting for this flexibility that the true value of on-orbit servicing can be evaluated. This chapter is organized as follows: Section 2 provides a background on on-orbit servicing, definitions and taxonomy, and includes a brief historical perspective of on-orbit servicing missions. Section 3 and 4 present a brief literature review of several on-orbit servicing studies and discuss the limitations of the traditional approach to on-orbit servicing. Section 5 proposes a new perspective on on-orbit servicing where the problem is analyzed from the servicing **customer's perspective**, instead of the usual (servicing) **provider's perspective**. The focus in Section 5 is on the **value** of servicing, and is studied independently from its cost. Advantages and limitations of this new approach to on-orbit servicing are also explored. Section 6 concludes the findings and discussion.

6.2 On-Orbit Servicing: Background

6.2.1 Definition and Taxonomy

On-orbit servicing comprises space assembly, maintenance, and servicing tasks to enhance the operational life and capabilities of space assets. Waltz (1993) describes these three functions of on-orbit servicing in the following terms:

> **Assembly** is the fitting together of manufactured parts into a structure, a subsystem, or elements of a subsystem. It is the on-orbit joining or construction of space systems and includes the deployment of solar arrays, antennas, and other appendages into their operational their operational configurations... [Assembly] occurs before a space system becomes [fully] operational.

> **Maintenance** is the upkeep of facilities or facilities or equipments [in space] either as necessitated or as directed by a scheduled program...**Preventive maintenance** includes observation, inspection, surface restoration, realignment, recalibration, repair, replacement of modules, contamination removal, test and checkout. **Corrective maintenance** includes all actions performed as a result of a system failure.

> **Servicing** includes the on-orbit replenishments of consumables and expendables... [However] the word *servicing* is often used to depict any or all of the functions named above.

Lamassoure (2001) provides a different taxonomy of on-orbit servicing, as seen from the *customer's perspective*, instead of the traditional classification based on the on-orbit servicing *provider's perspective*. This classification consists of:

> **Life extension** includes any on-orbit activity aimed at extending the operational life of the system in its initial design. This involves refueling, refurbishing and repairing.

> **Upgrade** includes any on-orbit activity aimed at improving the operational system in meeting its original mission goals.

> **Modification** includes any on-orbit activity performed in order to make a space system meet new mission goals. Examples include design changes through payload addition.

In addition to either of the above classifications, another important partition of on-orbit servicing concerns the timing nature of the servicing activity; it can occur **on-demand** or on a **scheduled** basis.
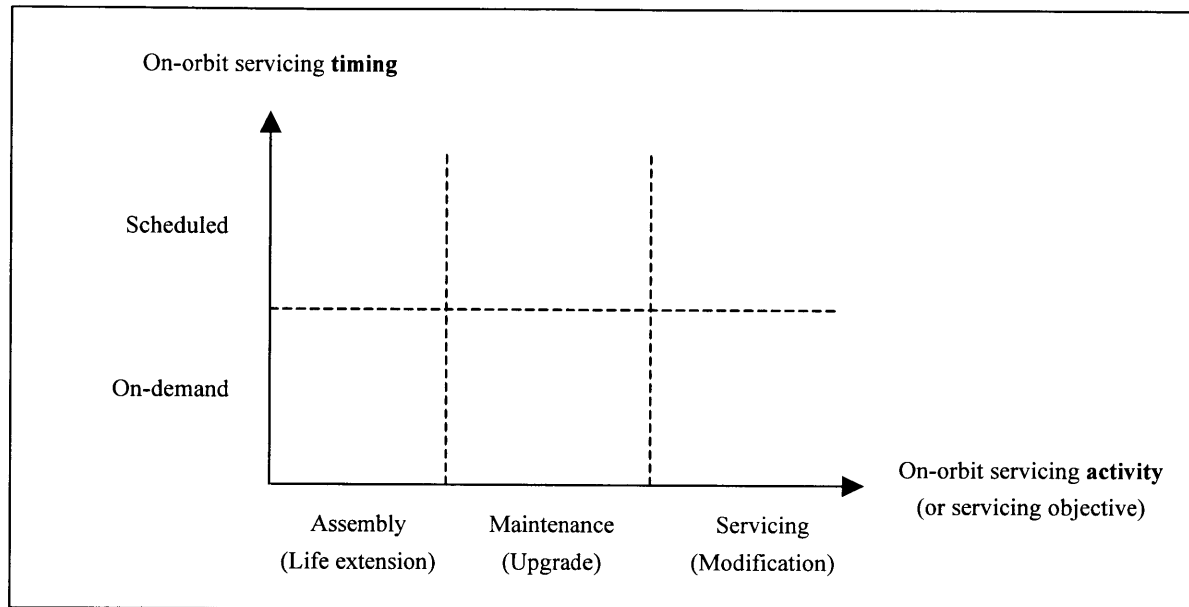
Fig. 6-3. On-orbit servicing taxonomy: Timing versus on-orbit servicing activity (or servicing objective).

Reynerson (1999) introduced a cost consideration in defining on-orbit servicing and *serviceable spacecraft*. Since any spacecraft can be serviced on-orbit given infinite resources, a spacecraft should not be considered *serviceable* unless the cost of servicing is justified by the benefits of doing so. His definition of a *serviceable spacecraft* follows from this reasoning:

> **Serviceable spacecraft**: Any spacecraft for which the benefits of on-orbit servicing outweigh the associated cost. The purpose of servicing can be to replace failed or degraded components, to upgrade existing capabilities, or to add new functionality or capability.

6.2.2 Historical Perspective

Although on-orbit servicing became largely known through the Hubble Space Telescope experience, it has nevertheless been practiced since the early years of human space flight. Waltz (1993) discusses significant servicing events. These include:

149

- The Skylab servicing missions;

- The capture and repair in space of the Solar Maximum Mission (SMM) spacecraft;

- The on-orbit retrieval, repair, and redeployment of the SYNCOM-IV satellite;

- The on-orbit retrieval, attachment of a booster stage, and re-launching of the Intelsat 6 communication satellite;

- The Hubble Space Telescope repair and upgrade servicing missions; and many others

In this subsection, we will briefly discuss the on-orbit servicing of Skylab, Solar Maximum Mission, and the Hubble Space Telescope. The reader interested in a thorough discussion of the history of on-orbit servicing is referred to Waltz (1998) or the Spacecraft Modular Architecture Design Study (1996).

Skylab: Skylab was the United States' first experimental space station and solar observatory. It was launched into orbit by a Saturn V booster on May 14th, 1973, and plunged back into Earth on July 11th, 1979 scattering debris over the Indian Ocean and Western Australia. Skylab was discontinuously inhabited from May 25th, 1973 till February 8th, 1974.

The Skylab missions (SL-2, SL-3 and SL-4) included scheduled maintenance activities, but also experienced immediately after liftoff (SL-1, unmanned), severe technical problems that required major unplanned maintenance efforts. Immediately after lift-off, the meteoroid shield, designed also to shade Skylab's workshop, deployed inadvertently and was torn away from the space station by atmospheric drag. One of the two solar panels of the craft was ripped off, and a strap of debris from the meteoroid shield wrapped around the other solar panel preventing it from deploying. This event and its effects prompted NASA, in an intensive 10-day period, to improvise new procedures and train the crew to perform unplanned extravehicular activity (EVA) in order to make the station operational and habitable. The various maintenance and repair activities performed by the successive crew included [Waltz, 93]:

- Installation and deployment of a solar shield "parasol" that cooled the inside of the overheating station from 52°C to 24°C

- Release and deployment of the jammed solar array

- Installation of a rate gyro package

- Major microwave antenna repairs, and coolant system maintenance

150

Skylab was NASA's first experience with on-orbit servicing. It demonstrated the effectiveness of crew members to perform complex and unplanned repair tasks, without which Skylab would have been doomed to failure immediately after launch, and the 3400-hour of on-board scientific experiments (solar observation, Earth observation, Biomedical investigations, etc.) would not have occurred. This raises the question of the value of on-orbit servicing versus its cost and the risk associated with performing it. In the case of Skylab, the value of salvaging the station and maintaining it habitable for its eight-month mission was regarded as sufficiently high to outweigh the cost and risk of servicing the station.

The Solar Maximum Mission (SMM): The Solar Maximum Mission was designed to provide coordinated observations of solar activity, in particular solar flares, particle acceleration, formation of hot plasma, and mass ejection, during a period of maximum solar activity. The spacecraft was launched on February 14th, 1980 into a quasi-circular orbit (512km–508km; inclination = 28.5°). Initially **designed for a two-year mission** [Adams et al., 87], the 2315kg spacecraft experienced after ten months of operations a failure in its Attitude Control Subsystem (ACS) that prevented the spacecraft from accurately pointing its instruments at specific regions in the Sun. In addition, one instrument, the coronagraph/polarimeter, showed pronounced deterioration in its performance. The problem was traced back to its main electronics box (MEB). Following the failure of the three momentum wheels, the spacecraft was put in back-up slow-spin mode, thus allowing the spacecraft to collect sufficient energy on its solar panels, but precluding the use of three instruments. In other words, the failure of the ACS and the corrective action taken to salvage the mission (spin mode) dramatically crippled the spacecraft's ability to meet its scientific objectives.

A repair mission was decided to prove the Space Shuttle's capabilities to rendezvous, repair, check out, and redeploy a free flying spacecraft (SMM was the first unmanned spacecraft to be serviced). In April 1984, after a year-long training at various NASA facilities, astronauts on-board STS-41C[1] (Challenger) captured the spinning spacecraft, replaced its attitude control module (primary objective), and repaired the faulty main electronics box of the coronagraph (secondary objective). SMM was then checked out, released into space, and resumed full operation. **The SMM repair mission extended the lifetime** of the spacecraft from two years to an additional five years after the repair, thus allowing for better coverage of the solar activity

---

[1] The flight also deployed the huge Long Duration Exposure Facility (LDEF).

cycle. SMM collected data until November 24th, 1989 and re-entered Earth's atmosphere on December 2nd, 1989.

NASA estimated that a successful repair mission of the SMM would restore the $230 million spacecraft at one-fourth of its replacement cost [Adams et al., 87]. Indeed, the cost of the repair mission was estimated by Goddard Space Flight Center (GSFC) at $60 million[2] [Waltz, 93]! Consequently, it was considered cost-effective to opt for on-orbit repair of the SMM over total spacecraft replacement.

Astronauts versus tele-robots/operators: The most difficult task executed by the astronauts, the main electronic box (MEB) repair sequence, was later performed by a 7 degree-of-freedom, force reflecting, controller/effector manipulator system on a full-scale mock-up of the Solar Maximum spacecraft. This technology demonstration was successfully completed at GSFC robotics laboratory between February 26th, 1987 and March 4th, 1987. It was attended by 400 NASA engineers, scientists, astronauts and government officials, and is documented in [Adams et al.,87].

Servicing the Hubble Space Telescope: NASA's Hubble Space Telescope is the first observatory designed for routine maintenance, upgrade, and refurbishment on orbit. The program is a 15-year mission with scheduled service by Shuttle astronauts every three years. Hubble's modular design allows for more than 90 spacecraft components and all of the scientific instruments to be replaced on orbit. Servicing maintains the spacecraft and allows for incorporation of new technologies.

Hubble was launched on April 24, 1990 with a full component of six scientific instruments. At that time, three new scientific instruments were already planned and an inventory of spare HST hardware had been acquired under the initial development contracts. HST budgets were sized to develop new instruments, maintain the spare hardware, sustain hardware expertise, plan and develop servicing activities, and test and integrate the payloads with the Shuttle. The primary objectives of the Second Servicing Mission were:

---

[2] In retrospect, there seems to be at least an order of magnitude difference between the estimated cost of the SMM repair mission (Shuttle flight, astronauts' training, etc.) and its actual cost. This dramatic under costing can be partially accounted for if one remembers NASA's lobbying efforts in the early 80's to make the Space Shuttle the main launch vehicle at the detriment of all expandable launchers.

1. To install two new scientific instruments, the Near Infrared Camera and Multi-Object Spectrometer (NICMOS) and the Space Telescope Imaging Spectrograph (STIS);
2. To replace a degraded Fine Guidance Sensor (FGS) with an upgraded spare;
3. To replace two failing tape recorders, one with a spare and the other with a state-of-the-art Solid State Recorder (SSR).

Development cost for the two scientific instruments are estimated at $105M for NICMOS and $125M for STIS. The upgrade to the FGS cost $8M and the balance of the hardware, including tools comes to $35M. Associated ground activities in support of the mission include new software and operations procedures development and testing, and mission planning and training, and cost $74M. The accomplishment of these objectives expanded and improved on the observatory's scientific capability and efficiency. NICMOS expanded Hubble's observing range to infrared light. STIS replaced the two spectrographs from the original payload, providing more efficient spectroscopy and discovery potential. The FGS is part of the pointing control system for the observatory and is also used for scientific observations. The spare FGS replaced a unit that was degrading and predicted to fail before 1999 (the next scheduled servicing). The upgrades to the replacement FGS increased pointing efficiency and reliability and increased the scientific potential of the telescope. The new Solid State Recorder has 10 times the storage capacity of the old tape recorders and because it is solid state, it has no moving parts to wear out.

Table 6-1 Hubble Space Telescope (HST) servicing cost break-down

HST Programs & STS-82 Costs ($m)	
NICMOS	105
STIS	125
FGS	8
Other flight hardware	35
Simulators/Testing	46
Ops/Software Development	28
HST Servicing Costs	347
Nominal Shuttle Flight Costs	448
Total Servicing Cost	795

Replacing Hubble's Main Computer: Hubble's main computer is responsible for monitoring the health of its many systems, for controlling the movement of the telescope from target to target, and for holding the telescope steady when observing. The computer, called the DF-224,

was designed in the late 1970's and its capabilities are much less than today's modern computers. Programming requires very specialized skills, unique to this computer, and maintaining the software is difficult and expensive. The DF-224 computer has degraded over time and during the First Servicing Mission in 1993 it was augmented with an additional computer called a co-processor. The design of the co-processor was based on the Intel 80386 microchip. During the third servicing mission, astronauts replaced this DF-224/coprocessor combination with a completely new computer based on the Intel 80486 microchip. This new computer is 20 times faster, and have six times as much memory, as the current computer on Hubble. The greater capabilities of the new computer increased productivity for the Hubble observatory by performing more work in space and less work by people on the ground. The result is decreased cost for software maintenance.

6.3 On-Orbit Servicing: A Brief Literature Review

While NASA engineers and astronauts were occasionally designing, training for, and performing on-orbit servicing, other members of the space community were investigating the design and consequences of a space-based servicing infrastructure. Indeed numerous studies have been published since the early 1980's addressing various issues related to on-orbit servicing, such as:

- The analysis and design of on-orbit **servicing architectures** (e.g., [SAMS, 88], [Leisman *et* al., 99])

- The identification of **serviceability requirements**, and spacecraft design implications (ability of a satellite to be serviced) (e.g., [SAMS, 88], [Hall and Papadopoulos, 99], [AIAA-G-042-1991])

- The design of **robotic on-orbit servicers,** and the identification of technical challenges associated with performing on-orbit servicing (ability of a "host vehicle" to provide servicing) (e.g., [Cook and Lindell, 99], [Kerstein *et* al., 94], [Matunaga, *et* al., 96], [Polites, 99], [Reynerson, 99])

- The **cost/benefit analysis** of on-orbit servicing (e.g., [SAMS, 88], [Leisman *et* al., 99], [Davinic *et* al., 97])

The Space Assembly, Maintenance, and Servicing (SAMS) study: The SAMS study is the most extensive study of on-orbit servicing in the literature. It was performed in 1986-1987 by two contractors headed by TRW Space and Technology Group and the Lockheed Missiles and Space Company. The program was a joint effort between the Department of the Air Force, the Strategic

154

Defense Initiative Office (SDIO), and NASA. The study sponsors provided the contractors with five design reference missions (DRM) as a means of exercising the SAMS study process for realistic conditions. From these DRMs, program requirements were generated and scenarios written for the spacecraft to be serviced, the hardware/tools to do the servicing tasks, and the space/ground infrastructure to support a SAMS program. The SAMS architecture that was developed included [Waltz, 93]:

- Servicing facilities at the Space Station (Freedom at that time)
- A reusable orbital transfer vehicle using cryogenic propellants
- A remotely piloted orbital maneuvering vehicle (OMV), which can carry a servicing front end and appropriate spare modules for the serviced satellite
- A facility for the on-orbit storage and handling of cryogenic propellants
- A propellant transfer system, which can service satellites with storable propellant
- A tele-operated satellite servicer system, with dual servicing arms and stowage for fuel
- A manned orbital transfer module, which can be carried to a remote servicing location

The study assumed routine and cheap access to space, and was dependent to a large extent on the presence and support of humans in space. The seven-year program however was terminated after 16 months (Phase I). Its scope, (grand) scale, and assumptions proved its downfall. The study failed to inspire confidence in its conclusions regarding the cost-effectiveness of on-orbit servicing.

After the SAMS study, the focus of on-orbit servicing studies shifted from high-cost manned servicing infrastructure, to unmanned low cost robotic servicing missions with the potential to reduce life cycle costs of high value space systems. The spacecraft modular architecture design study (SMARD) and the on-orbit servicing of the GPS constellation study illustrate this trend. These two studies are summarized below.

The Spacecraft Modular ARchitecture Design (SMARD) study: The focus of the SMARD study was on unmanned low cost robotic servicing missions that have the potential to enhance the performance or reduce the life cycle cost of high-value on-orbit assets. The study was performed in 1996 by the Naval Research Laboratory, and is documented in [Davinic *et* al., 97] and [Reynerson, 99]. The study first identified and categorized different levels of servicing for a

remote sensing constellation[3]. Components of the satellite architecture were examined to determine the potential for replacement by a servicing mission: It was shown that one third of the satellite components can be practically replaced, and many more could be replaced by adopting a more modular bus and payload design. Design modifications were suggested to make satellites better apt to being serviced. The study determined the following set of servicing needs of the satellite system:

- Replenishment of consumables and degradables (propellant, batteries, solar array)
- Replacement of failed functionality (payload and bus electronics, and mechanical components)
- Enhancement of mission through insertion of new technology

On-orbit replacement of components in the SMARD study is performed functionally, not physically: All the replacement components are packed in a single payload module, which a servicer satellite attaches to a docking interface on a satellite. This functional replacement strategy is considered to minimize cost and complexity of the servicing mission, and is in contrast with physical replacement strategies advocated by other on-orbit servicing studies that consider human or robotic manipulation and (physical) replacement of failed or degraded hardware. Electrical and mechanical considerations were addressed to allow for functional replacement of components (modular data architecture design, docking interface, etc).

A point design solution for a satellite servicer was developed as part of the study. The servicer consists of two payload modules and one bus module. Each payload module contains replacement components for one satellite. A servicer can thus repair or upgrade two satellites. "The point design was developed in such detail that a credible bottoms-up costing analysis could be conducted" [Reynerson, 99]. A costing evaluation was performed to determine the impact of servicing on the life-cycle cost of the constellation. The evaluation had three distinct components:

1. A cost evaluation of the proposed servicer vehicle was conducted. The costing included all design, development, integration, and ground test efforts.
2. An estimate was made of the cost impacts associated with redesigning the current satellites in the constellation to make them serviceable.
3. A set of lifecycle costs was developed for several on-orbit scenarios.

---

[3] The architecture consisted of 10 satellites in Low Earth Orbit, with two satellites per plane. Details of the constellation and mission are considered classified.

The study reports lifecycle cost savings from 10.3–38.2%, depending on the targeted life extension (from two to six years) and the number of servicers used, over a period of 20 years.

Despite its credible technical details and its encouraging cost-benefit analysis, the SMARD study did not have a follow-up. The advantages shown in the study in terms of cost savings and availability did not outweigh the perceived technological risk and cost uncertainty associated with performing on-orbit servicing.

The On-Orbit Servicing of the GPS Constellation study: Two companion studies performed at the Air Force Institute of Technology [Leisman et al., 99] and the Aerospace Corporation [Hall and Papadopoulos, 99] addressed the problem of servicing the GPS constellation. Leisman et al. (1999) evaluated multiple architectures for on-orbit servicing of the GPS constellation, and explored the costs and benefits of upgrading/repairing GPS satellites through Robotic Servicing Systems (RSS). Their study, however, did not address "the complex technical and contractual modifications that would be necessary to make GPS satellites serviceable": The structural modifications necessary to enable the servicing of the GPS IIF spacecraft were addressed by Hall and Papadopoulos (1999).

The objectives of the first study [Leisman et al., 99] were to identify the logistical support needs of the GPS constellation, to find multiple servicing support solutions, and to identify which of these solutions best meet those needs. The study proceeded as follows:

First it identified logistical support needs of GPS constellation through interviews with GPS managers, and mapped the criteria decision makers consider important in evaluating a Robotic Servicer Satellite (RSS). Responsive upgrade of the GPS constellation turned out to be of primary concern to GPS managers, while repair was considered desirable but not necessary.

> New technology or capabilities are provided only as the current satellites retire. The next generation of block IIF will have a design life of 12.7 years. Thus in the future, providing the full constellation with new capabilities will require [...] approximately 13 years. The problem to be solved in this study is how to decrease cycle time for implementing new capabilities while still minimizing costs.

Second, the study defined multiple architectures that could best meet customer needs. Architectures were differentiated according to the number of robotic servicers (RS) used per orbital plane, the type of propulsion system adopted, and the mass delivery capacity (orbital replacement units -ORU- of 50kg, 150kg, and 300kg).

Third, each architecture was evaluated for costs and benefits over a 15-year operational period, and four servicing missions to each satellite. Costs were estimated using the NASA/Air Force NAFCOM 1996 parametric cost analysis program.

Finally, the study concluded that on-orbit servicing of the GPS constellation offers greater benefits and would be less costly than the current GPS satellite management paradigm (current policy of 2 satellite replacements/year):

> Using current methods, the average cost of replacing a GPS satellite is approximately $100 million. The most expensive of the top six [on-orbit servicing architectures] could upgrade the entire constellation for $60 million per satellite.

Hall and Papadopoulos (1999) complemented the previous study by conducting a preliminary assessment of structural modifications necessary to make the GPS spacecraft serviceable. The study focused on satellite upgrade through the addition of new components. Design modifications included upgrade slots that would be added to the GPS satellite baseline design and launched empty. The authors used (and modified) mass estimate relationships to evaluate the additional required to make the spacecraft serviceable. For instance:

> Additional thermal control mass was added to account for increased complexity in thermal interfaces and heat loads that are added on-orbit. Instead of the baseline 3.7% of dry mass [mass of thermal control subsystem], 4–7.5% was used.

The study concluded that an additional mass of 3–15% would be needed in order to render the GPS spacecraft serviceable (baseline wet on-orbit mass of 2813lb/1280kg). The study, however, did not address design modifications at the subsystem level. This omission on one hand, degrades the accuracy of the result, and on the other hand, fails to show whether the on-orbit servicing of the GPS constellation is actually feasible (even though the companion study showed that it was economical).

6.4 Limitations of the Traditional Approach to On-Orbit Servicing

The studies discussed above represent typical samples of the traditional approach to on-orbit servicing. To some minor variations, they all proceed as follows: First, the levels of logistical support for a given space mission are identified. Then, on-orbit servicing architectures are proposed that could meet these serviceability requirements. Parallel to this phase, designs for host vehicles that could perform on-orbit servicing are proposed, and design modification necessary to make spacecraft serviceable are addressed. Finally, the cost-effectiveness of on-orbit servicing is assessed. This process is schematically depicted in Figure 6-4.

On-orbit servicing architecture	Modifications required to make spacecraft serviceable	Design of host vehicle to perform on-orbit servicing
# servicers per plane per satellite	Component accessibility	Autonomous Rendezvous and Docking capability (AR&D)
Characteristics of ORU canisters	Docking adapter/interface	Dexterous manipulator(s)
Space depots	Modularity, standardization (e.g., HST modular design allows 90 components to be replaced)	Electrical/mechanical interface

Cost-effective?

Fig. 6-4. Sequence of issues addressed in the traditional approach to on-orbit servicing. Cost-effectiveness of on-orbit servicing is left as an output of such studies.

While this traditional approach to on-orbit servicing often represents sound systems engineering practice, and offers numerous advantages (e.g., addressing the technical feasibility of on-orbit servicing), it nevertheless has intrinsic limitations that hamper the ability to make meaningful conclusions regarding the cost-effectiveness of on-orbit servicing. These limitations are discussed below.

Cost Estimate Relationships are inappropriate to estimate the cost of a robotic servicer: Spacecraft costs depend on their size, complexity, technology readiness (TRL), design lifetime, as well as other characteristics. Several governmental organizations have developed over the years Cost Estimate Relationships (CERs) that relate spacecraft cost, or subsystem cost, to physical, technical, and/or performance parameters. The CERs are based on an appropriate historical database of past satellites programs. The basic assumption of parametric cost modeling is that

"satellites will cost next time what they cost the previous time." Thus the use of CERs to estimate the cost of a robotic servicer is doubtful since a servicer satellite would be substantially different from the historical data that was used to establish the CERs.

On-orbit servicing cost advantages remain smaller than cost uncertainty: Assuming the error in using CERs to estimate the cost of a robotic servicer can be quantified, Lamassoure (2001) showed that while there are situations in which on-orbit servicing proves cost-effective, the cost advantage of on-orbit servicing remains smaller than the cost uncertainty, thus making "any definitive conclusion about the cost-effectiveness of servicing impossible." Figure 6-5 illustrates this point by comparing the probability distribution function of three different cost models for a typical servicer with 200kg of cargo/payload.
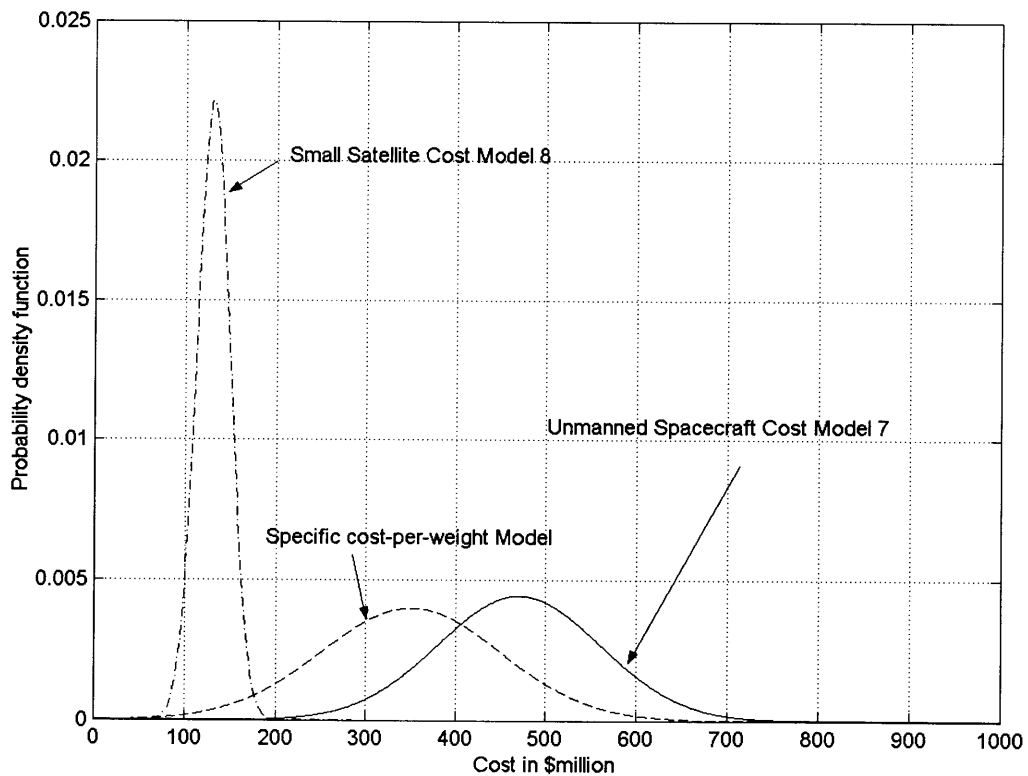
Fig. 6-5. Comparison of cost models results for 4 servicers with 200kg cargo/payload. Adapted from Lamassoure (2001).

The price a spacecraft would pay for being serviced is not necessarily equal to the servicing cost: In the traditional approach, the cost of the servicing architecture was compared with the overall constellation lifecycle cost savings in order to assess the cost-effectiveness of on-orbit

servicing (all the previous studies that have addressed the cost-effectiveness of on-orbit servicing have developed servicing architectures for a **constellation** of satellites, e.g., GPS, classified LEO constellation in the case of the SMARD study). However, the price a spacecraft would pay to be serviced also depends on the development policy for the servicing infrastructure, and it is not reasonable to assume that the cost of a servicing architecture would be amortized by a single spacecraft and over a single servicing event. The cost of the whole servicing infrastructure can be amortized over several missions, or can be borne by a government agency such that only the marginal cost of servicing would be charged to individual spacecraft [Hastings *et* al., 01]. This undermines the traditional strategy of investigating the cost-effectiveness of on-orbit servicing.

The traditional approach to on-orbit servicing overlooks the intrinsic value of servicing for a space mission: Traditionally, on-orbit servicing has been analyzed from the (servicing) provider's point of view. It is surprising that no previous study has incorporated the (potential) customer's perspective on the subject. The value of servicing for a space mission should exist independently of any servicing architecture. In addition, by using traditional valuation tools such as net present value (NPV) calculations, previous studies have underestimated an important component of servicing value: Servicing provides space missions with options to react to the resolution of uncertain parameters (e.g., evolving market needs, changing military contingencies). This **flexibility** is a significant advantage of servicing, however its value is not captured by NPV calculations [Lamassoure, 2001]. Decision Tree Analysis (DTA) and Real Options calculations are more appropriate tools to capture the flexibility component in the value of servicing.

It is difficult to make a convincing case of the cost-effectiveness of on-orbit servicing given the intrinsic limitations of the traditional approach discussed above. This motivates the development of a new perspective on on-orbit servicing that includes the (potential) customer's perspective, and where the value of servicing, including the value of flexibility it provides, is studied independently of the cost of servicing. This is elaborated in the following section.

6.5 A New Perspective on On-Orbit Servicing

The traditional approach to on-orbit servicing fails to recognize the intrinsic value of servicing for a space mission. This value, which we will define for the time being as the maximum price a space mission would be willing to pay for the on-orbit asset to be serviced, should exist independently of any servicing infrastructure. Highlighting the **value** of servicing adds a new dimension to on-orbit servicing studies, and shifts the focus from the traditional (servicing) provider's perspective to the (potential) customer's perspective. Figure 6-6 illustrates the two stakeholders' perspectives on on-orbit servicing.
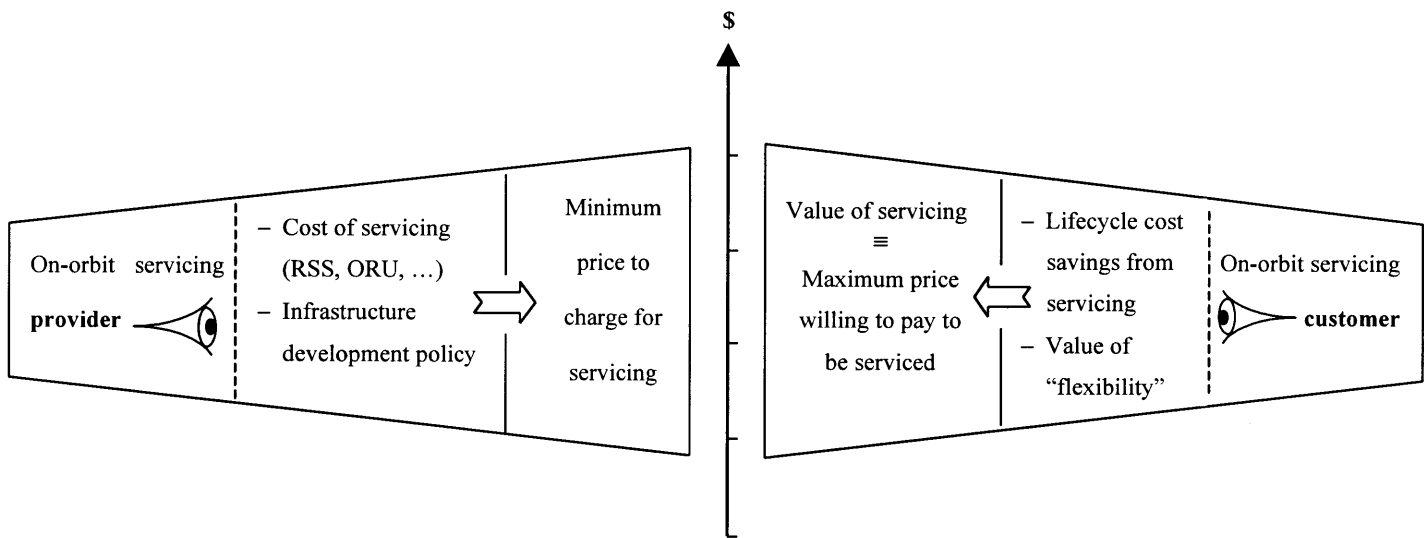
Fig. 6-6. On-orbit servicing provider's perspective versus customer's perspective.

The traditional approach to on-orbit servicing has explored (parts of) the left segment of Figure 6-6 (see Fig. 6-4 for more details on this segment). Suggestions have been made to investigate the effect of a servicing development policy where the cost of a servicing infrastructure would be borne by a government agency, and only the marginal cost of servicing charged to individual spacecraft. Despite these efforts, not much confidence was shown in the traditional approach conclusions regarding the cost-effectiveness of on-orbit servicing (for reasons discussed above). Ultimately, the decision to service an on-orbit asset lies with the potential customer (customer-centric perspective): A potential customer would opt for servicing if the value of servicing (V_{serv}) his/her spacecraft exceeds the cost to service it, or the minimum price a provider can afford to charge for servicing (P_{min_serv}), given a servicing architecture, an infrastructure development policy, etc.). This observation is captured in Eq. (6-1).

162

$$V_{serv} \equiv \underbrace{P_{max\_serv}}_{\substack{\text{as seen from the} \\ \text{customer's perspective}}} > \underbrace{P_{min\_serv}}_{\substack{\text{determined by} \\ \text{the provider}}} \qquad (6\text{-}1)$$

Where:

$$\begin{cases} P_{min\_serv} = P(\text{servicing architecture, design of servicer, infrastructure development policy, etc.}) \\ \\ P_{max\_serv} = P(\text{lifecycle cost saving, value of flexibility, etc.}) \end{cases}$$

Separating the value of servicing from its cost presents several major advantages. First, the conclusions drawn are not dependent on a particular servicing architecture; instead they reflect the potential customer's valuation of on-orbit servicing independently of any servicing solution. Second, separating the value of servicing from its cost significantly reduces the uncertainty in the results that plagues the traditional approach to on-orbit servicing. Third, in identifying the maximum cost cap below which servicing makes economical sense, this approach helps guide the selection of space missions to target for servicing, and provides a justification for a development policy of a servicing infrastructure. In addition, a major component of the value of servicing, the value of flexibility on-orbit servicing provides to space missions, is not taken into account by the traditional approach. Indeed, on-orbit servicing provides decision makers with **options** (to refuel, repair, upgrade, modify) that don't need to be set prior to launch. Instead, the decision to **exercise** such options depends on the resolution of parameters that were uncertain at the time of launch (e.g., market demand/uncertainty, military contingency, etc.). The value of this flexibility is not captured by standard discounted cash flow techniques such as the Net Present Value (NPV) or the Internal Rate of Return (IRR) used by previous studies of on-orbit servicing. In the following, we argue that only by accounting for this flexibility can the true value of on-orbit servicing be captured.

6.5.1 Accounting for Flexibility Provided by On-Orbit Servicing

The new perspective on on-orbit servicing presented herein is based on three main ideas. The principal idea of this new approach consists of estimating the **value of servicing** separately from its cost, thus shifting the focus from the traditional (servicing) provider's perspective to the

(potential) customer's perspective. The second idea lies in the observation that on-orbit **servicing provides flexibility** to space missions, as discussed previously. And finally, contrary to what has been implicitly assumed by traditional approaches, the value of servicing is not limited to potential cost savings; instead the **value of flexibility** provided by on-orbit servicing represents an important component of the value of servicing. In other words, the third idea consists in recognizing that the value of servicing should account for the value of flexibility provided by on-orbit servicing. Traditional discounted cash flow techniques such as the standard NPV calculation used by previous studies of on-orbit servicing cannot capture the value of flexibility. Decision-Tree Analysis on the other hand is a more elaborate capital budgeting tool that is capable of accounting for the value of flexibility, and is particularly useful for analyzing complex sequential decisions, and in situations where uncertainty is resolved at distinct, discrete points in time. This is further discussed in the following section.

6.5.2 Failure of Traditional Valuation Tools to Capture the Value of Flexibility: Example of a Standard NPV Calculation versus Decision-Tree Analysis

The following example, adapted from Lamassoure (2001), illustrates the shortcoming of the traditional NPV calculation to capture the value of flexibility, and contrast it with the use of Decision-Tree Analysis, a more elaborate capital budgeting tool than the NPV that is capable of accounting for the value of flexibility. A substantial body of literature exists that describes the shortcoming of NPV calculations; the reader is referred to Faulkner (1996), Trigeorgis (1996), or Amram and Kulatilaka (1999) for more details.

> *Assume a project has a current value S = $200m and its value after one year is discrete but uncertain: it can either increase to S^+ = $400m with a subjective probability p, or decrease to S^- = $100m. The owner of the project gives a potential buyer the option, but not the obligation, to acquire the project after one year for a price E = $280m. What is the value of this option? In other words what price for the option will the owner and potential buyer agree upon?*

For discrete cash inflow C_n and outflow I_n over N periods of time, with a risk-adjusted discount-rate k, the standard NPV calculation can be written as:

$$NPV = \sum_{n=1}^{N} \left\{ \frac{C_n}{(1+k)^n} - \frac{I_n}{(1+k)^n} \right\} \tag{6-2}$$

In our example, the NPV of buying the project is:

$$NPV = p\frac{S^+ - E}{1+k} + (1-p)\frac{S^- - E}{1+k} \qquad (6\text{-}3)$$

Assuming equal probability for the project value to go up or down, i.e., $p = 0.5$, and taking a risk-adjusted discount rate $k = 20\%$, we get:

$$NPV = -\$25m$$

So from an NPV perspective, the project is not interesting, and the option to acquire it at the conditions stated above will be discarded. This calculation however fails to take into account the managerial flexibility resulting from the asymmetry in having the right, but not the obligation, to acquire the project after one year. In order to avoid this deficiency of the traditional valuation, we revert to Decision-Tree analysis (DTA).

Decision-Tree Analysis is a particularly useful tool for analyzing complex sequential investment decisions, and in which uncertainty is resolved at distinct, discrete points in time such as in our example. DTA describes a sequence of decisions that are not set from the start, but depend on the resolution of some uncertain parameter(s). Unlike an NPV calculation, which is often misused by managers inclined to focus only on the initial decision to accept or reject a project at the detriment of subsequent decisions, DTA forces management to lay out an operating strategy, and to recognize explicitly the interdependencies between the initial decision and subsequent decisions [Trigeorgis, 96]. The optimal initial decision in a DTA is determined by starting from the end of the tree and working backward to the beginning. This dynamic programming, roll-back procedure involves determining at each stage the expected risk-adjusted discount NPV (or expected utility) by multiplying all NPV (or utility) values calculated at the previous–although chronologically following–stage with their respective probabilities of occurrences and summing up. Furthermore, the flexibility available to the decision-maker is taken into account by considering only optimal decisions made at each evolution of the value of the project. Let us see how this applies to our example.

Figure 6-7 is a simple decision tree representing our investment example. If the value of the project increases, the optimal decision for the potential buyer (holder of the option) is to exercise the option and thus acquire the project. The pay-off in this case is $\$(S^+ - E)$. If the value of the

project decreases, the optimal decision is not to exercise the option, i.e., not to acquire the project and thus avoid the losses. There is no pay-off if the project is not acquired.
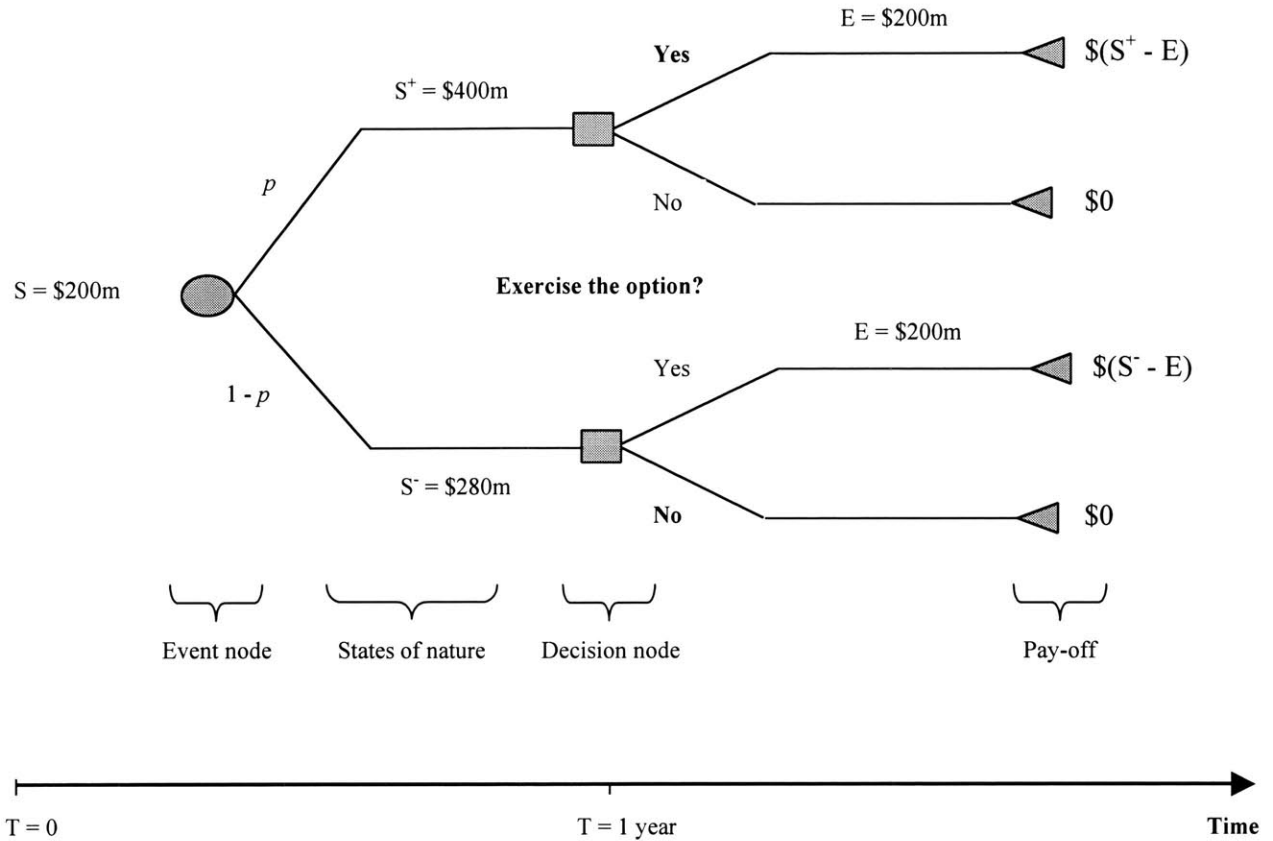
Fig. 6-7. Decision Tree representing the investment problem in our example (standard notation). Optimal decisions given each state of nature are written in bold.

The value of the option under these conditions becomes:

$$V_{DTA} = p \frac{\max(S^+ - E; 0)}{1+k} + (1-p) \frac{\max(S^- - E; 0)}{1+k} \tag{6-4}$$

Assuming equal probability for the project value to go up or down, i.e., $p = 0.5$, and taking a risk-adjusted discount rate $k = 20\%$, as in the previous calculations, we get:

$$V_{DTA} = \$50m$$

This shows that, assuming a rational decision-maker[4], the option of acquiring the project after one year is actually very attractive and is worth $50m. The difference between the NPV and the V_{DTA} results from the value of flexibility (V_{flx}) in having the right, but not the obligation to acquire the project after one year.

$$V_{flx} = V_{DTA} - NPV \qquad\qquad (6\text{-}5)$$

This simple example is used to illustrate two points: First, the standard NPV calculation used by previous studies of on-orbit servicing cannot capture the value of flexibility. Second, the value of flexibility can constitute a substantial part of the value of a (flexible) project. In other words, project valuation using standard discounted cash flow techniques, i.e., not accounting for flexibility when it exists, is erroneous and often dramatically underestimated.

Limitations of the Decision-Tree Analysis: Decision-Tree Analysis is one tool for capturing the value of flexibility. However, just like most tools, it has its limitations. First, it can often become an unmanageable "decision-bush analysis" when actually applied in realistic settings, as the number of different paths through the tree (or bush!) expands geometrically with the number of decisions, or states considered for each variable [Trigeorgis, 96]. Second, it can only account for a finite number of decision nodes, occurring at discrete decision times, following discrete variations of the unknown parameter(s). In other words, DTA cannot account for uncertain variables that are continuous. Third is the problem of determining the appropriate discount rate. Using a constant discount rate presumes the risk borne per period is constant; this is obviously not the case when options are available. Flexibility (availability of options) decreases a project's exposure to uncertainty, thus alters the project's risk. It is therefore more appropriate to use different discount rates in different periods. But the problem of finding the appropriate discount rate (per period or not) still remains. Option-Pricing Theory, and its spin-off, Real Option Theory are two other frameworks that capture the value of flexibility in financial and real assets, and that solve the problem of the discount rate. The application of Real Option Theory requires the identification of an appropriate underlying financial asset, or a "twin security" that has the same risk characteristics as the real asset (or the non-traded asset), in order to carry out the valuation. Such a twin security doesn't necessarily exist for some projects (or cannot be constructed), thus rendering a Real Option valuation impractical. While this is the subject of on-going research, it is nevertheless beyond the scope of this work. The reader is referred to Trigeorgis (1996) for an

---

[4] One that can make optimal decisions, i.e., that maximize pay-offs, after each decision node.

elaborate discussion of Real Option Theory, and to Neely and de Neufville (2000) for a discussion of the limitations and Real Options valuations and the development of a *Hybrid* Real Options framework. In this work, we will use Decision-Tree Analysis. While it represents an important improvement over traditional discounted cash flow techniques, and most importantly can capture the value of flexibility, its has nevertheless its limitations and would often undervalue a project when a constant discount rate is used throughout the tree.

So what are the options made available to space systems through on-orbit servicing? These are discussed in §6.2.1 and illustrated in Figure 6-3. They include the option to service a spacecraft for life extension, the option to upgrade a spacecraft, the option to modify its payload, and of course the option to repair after a random failure. The Hubble Space Telescope servicing missions are perfect examples of cases where all these options have been exercised (repair of its primary mirror, replacement of degraded Fine Guidance Sensor and failing tape recorders, upgrade of the main computer, and addition of two new scientific instruments).

In the following section, we will explore how potential customers of on-orbit servicing would assess the value of the flexibility (availability of options) provided by servicing, and discuss the implications of this valuation process.

6.6 Estimating the Value of Spacecraft Life Extension: Application of the New Perspective on On-Orbit Servicing to a Specific Instance of Flexibility

In the following, we apply this new perspective to capture the value of spacecraft lifetime extension provided by on-orbit servicing.

6.6.1 The Simple Case: Value of Servicing Through Minimizing Cost

In this case, we assume the customer, a non-profit organization for instance, seeks to evaluate three design alternatives, with the explicit purpose of achieving an effective lifetime of 15 years. We are not concerned in this example with a dynamical environment where issues of market uncertainty and technology obsolescence are relevant. The alternatives are the following:

i.   Launch a spacecraft designed for 15 years

ii.   Launch a spacecraft designed for T_0 years; After T_0, replace the spacecraft with another spacecraft designed for $(15 - T_0)$ years

iii.   Launch a spacecraft designed for T_0. After T_0, extend the lifetime of the spacecraft through on-orbit servicing (would include for instance refueling and/or replacing batteries, solar panels, thermal coating, etc.) to $(15 - T_0)$ years

Which alternative is the least costly for our customer?

Let us first explore alternatives (i) and (ii). We have recently investigated the effects of varying the spacecraft design lifetime requirement on various subsystems, and deduced spacecraft cost profile (and mass) as a function of this requirement, $C(T_{Life})$, all else being equal (see Chapter 4). A typical example of a spacecraft cost (to IOC) profile is given in Figure 6-8.

169
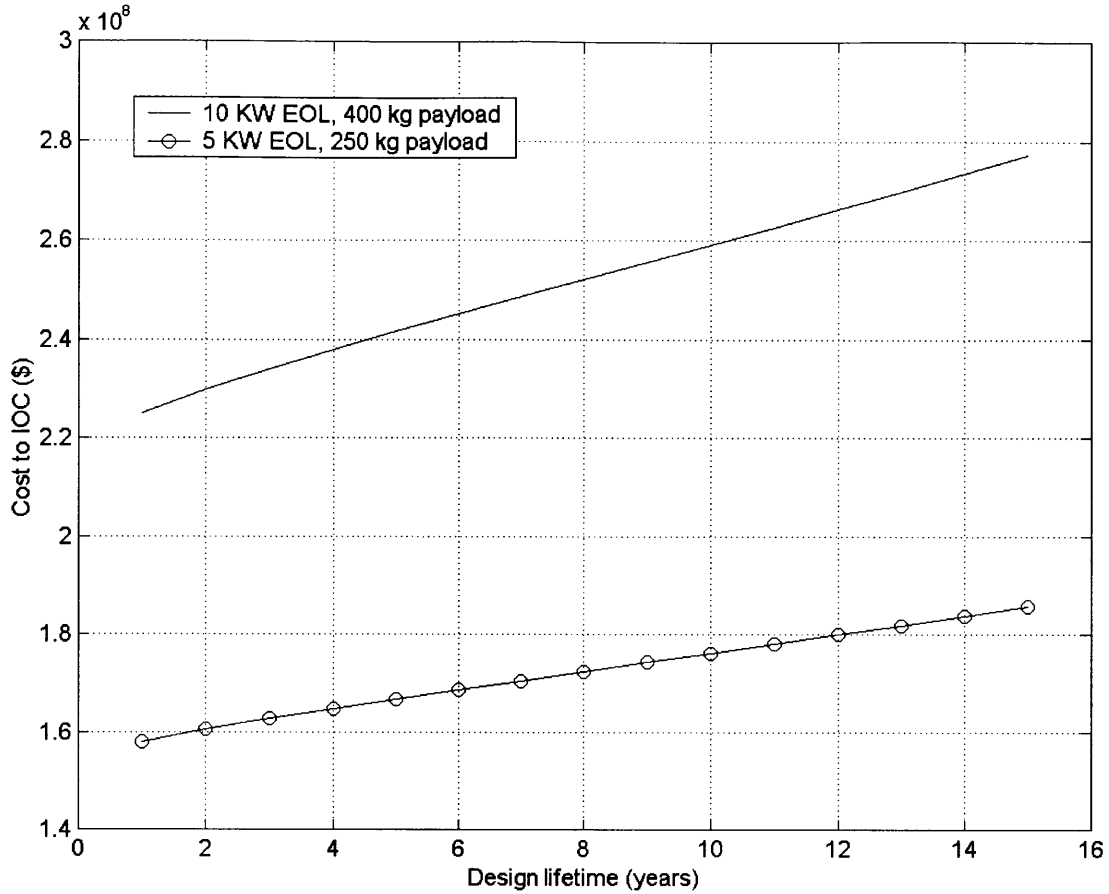
Fig. 6-8. Cost to IOC as a function of the design lifetime requirement (spacecraft in GEO, mission reliability = 95%, three-axis stabilized, GaAs cells, Ni-H2 batteries).

We define a quality factor for the staging of the spacecraft design lifetime as follows:

$$\rho(T_0, \Delta T) = \frac{C(T_0 + \Delta T)}{C(T_0) + C(\Delta T)} \tag{6-6}$$

ρ is the ratio of the cost of designing a spacecraft for $(T_0 + \Delta T)$, divided by the cost for designing two spacecraft for T_0 and ΔT respectively. For $\rho > 1$, it is less costly to stage the design lifetime in T_0 and ΔT than to design for $(T_0 + \Delta T)$ years. This illustrates the importance of establishing a cost profile, such as $C(L_{life})$, for all complex engineering systems, in order to guide the selection of the product's design lifetime requirement. Figure 6-9 shows a family of ρ for various T_0 and life extension ΔT.

5 KW EOL, 250kg Payload
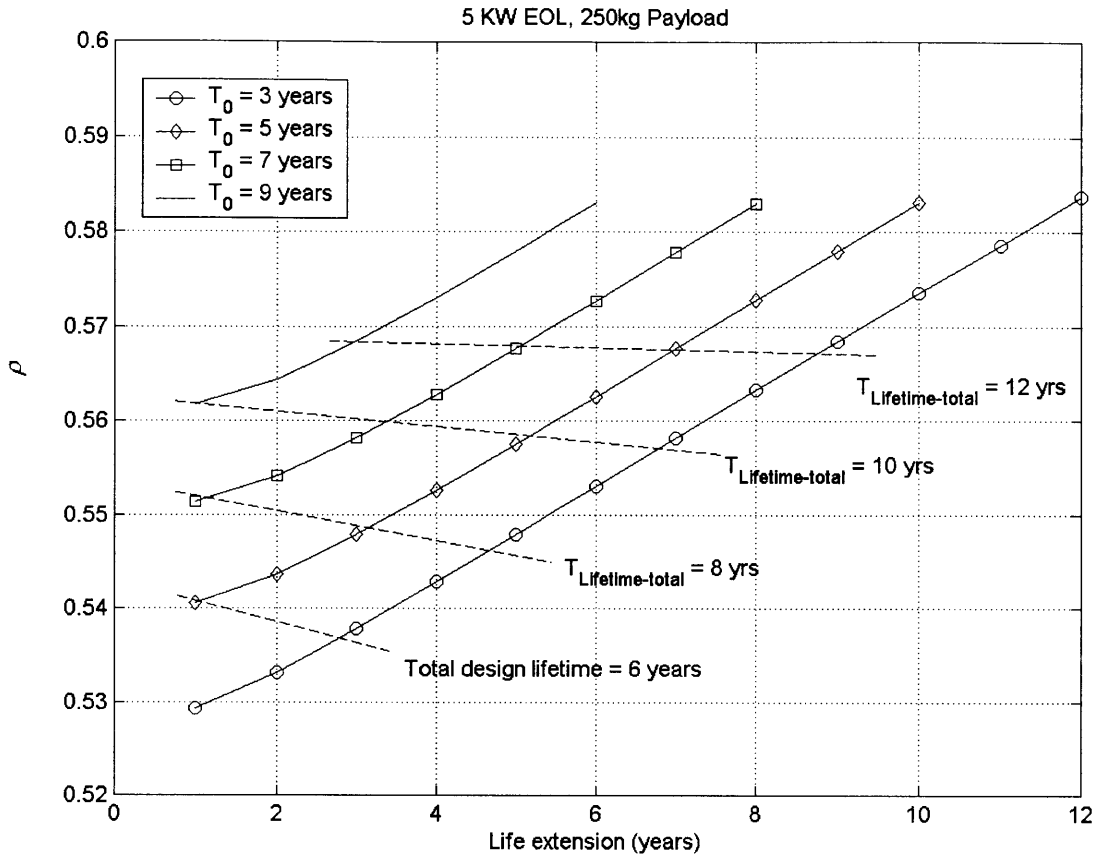
Fig. 6-9. Quality factor for the staging of the spacecraft design lifetime as a function of life extension.

A couple of observations are worth making based on Figure 6-9. First, we note that $\rho < 1$ for all T_0 and ΔT. In other words, it is always cheaper to design a spacecraft for the maximum required lifetime $T_{Life\text{-}total}$ than to stage the lifetime in two spacecraft designed for T_0 and $(T_{Life\text{-}total} - T_0)$. Second, for a given design lifetime $T_{Life\text{-}total}$, short life extensions are more expensive than longer life extensions (e.g., for $T_{Life\text{-}total} = 8$ years, it is more expensive to design two spacecraft for 7 years and 1 year, than two spacecraft for 5 years and 3 years). These conclusions are indeed expected given the high cost incurred to design and launch a spacecraft, and the smaller cost increments associated with increasing the design lifetime.

Alternative (i) is therefore always less costly than alternative (ii). What about alternatives (i) and (iii)? What is the maximum price the customer would be willing to pay to extent the design lifetime of his/her spacecraft through on-orbit servicing $(P_{serv\text{-}max})$, such that alternatives (i) and (iii) are cost-equivalent? This condition can be written as follows:

171

$$\underbrace{C(T_0 + \Delta T)}_{\text{alternative (i)}} = \underbrace{C(T_0) + P_{serv-max} \times e^{-rT_0} + (risk\ premium)}_{\text{alternative (iii)}} \qquad (6\text{-}7)$$

The left side of the equation is the cost to design a spacecraft for $(T_0 + \Delta T)$ years; it represents alternative (i). The right side represents the cost of designing a spacecraft for T_0 years, then extending its life for through on-orbit servicing. Since $P_{serv-max}$ is incurred at a later period than $C(T_0)$, i.e., T_0 years later, it is discounted accordingly (r: discount rate). In addition, because servicing involves tampering with a spacecraft, it is inherently riskier than alternative (i); a risk premium is thus added to the left side of the equation. Equation 6-7 can be written as follows:

$$P_{serv-max} = (1 - \Psi) \times [C(T_0 + \Delta T) - C(T_0)] \times e^{rT_0} \qquad (6\text{-}8)$$

$P_{serv-max}$ is the maximum price a customer would be willing to pay, after T_0 years, to extend his/her spacecraft design lifetime by ΔT, instead of designing it for $(T_0 + \Delta T)$ years from the start, such that alternatives (i) and (iii) are cost-equivalent. Ψ is an insurance premium contracted to mitigate the financial risk incurred due to the servicing operation; it is a decreasing function of the reliability of the servicing operation (as the probability of failure or crash into the host vehicle increases, Ψ obviously increases). Figure 6-10 shows a family of $P_{serv-max}$ for different design lifetimes and life extensions. One particular point on the plot reads as follows: The maximum price of servicing a customer would be willing to pay in order to extend the design lifetime of a spacecraft four additional years from seven to eleven years is approximately $17 million (with an insurance premium equal 20% of the cost savings from designing for 7 years instead of 11 years). If on-orbit servicing cannot be achieved within this cost cap, it is not cost-effective for the customer to have his/her spacecraft serviced for life extension
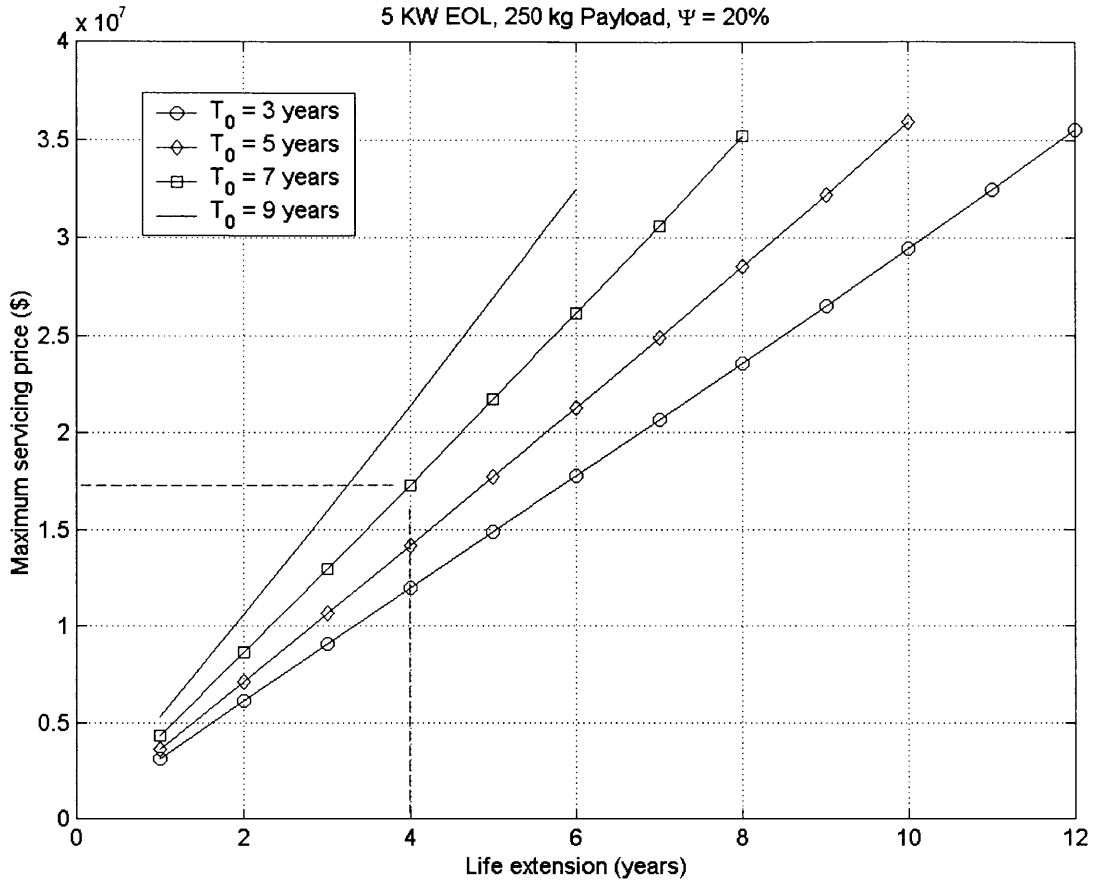
Fig. 6-10. Maximum servicing price as a function of life extension. A standard 10% discount rate is considered.

Figure 6-10 represents the maximum price a customer would be willing to pay to extent the design lifetime of his/her spacecraft through on-orbit servicing ($P_{serv-max}$), and for which alternative (i) and (iii) are cost-equivalent. These curves are solutions of Eq. 6-7; they represent the value of servicing for life extension as seen from the customer's perspective (see Fig. 6-6). As expected, the value of servicing increases as the lifetime extension increases (from $5m to $30m approximately). A potential customer would therefore opt for servicing only if the price charged for servicing is less or equal to the value of servicing ($P_{serv-max}$). Conversely, a servicing provider should constraint the design of a servicing architecture, robotic servicer, orbital replacement units (ORUs), etc. in order to be able to deliver the on-orbit service for less than $P_{serv-max}$, otherwise he/she will find no customer.

Let us further explore the idea of value of servicing through its impact on the spacecraft design lifetime. However instead of **life extension,** we consider on-orbit servicing as a mean to counter

spacecraft **life contraction** resulting from unanticipated but necessary orbit maneuvers. In order to do so, let us consider the following scenario:

A military communications satellite is designed for T_{Life} =10 years, with a 20% fuel margin for station keeping. The satellite was initially designed as part of a four-satellite constellation providing full Earth coverage. However due to a launch mishap, only three satellites are operational. Thus full Earth coverage is not achieved and one satellite has to perform phasing maneuvers in order to track changing contingency locations. Eq. 6-9 gives the incremental velocity ΔV required to change the satellite's phase by $\Delta\Phi$ in τ days:

$$
\begin{cases}
\dfrac{\Delta V_{ph}}{V_0} = 2\left|\sqrt{2-\left(\dfrac{\lambda}{\lambda-\Delta\Phi/2\pi}\right)^{2/3}}-1\right| \\[2em]
\lambda = Integer\left[\dfrac{\tau}{T_0}+\dfrac{\Delta\Phi}{2\pi}\right]
\end{cases}
\tag{6-9}
$$
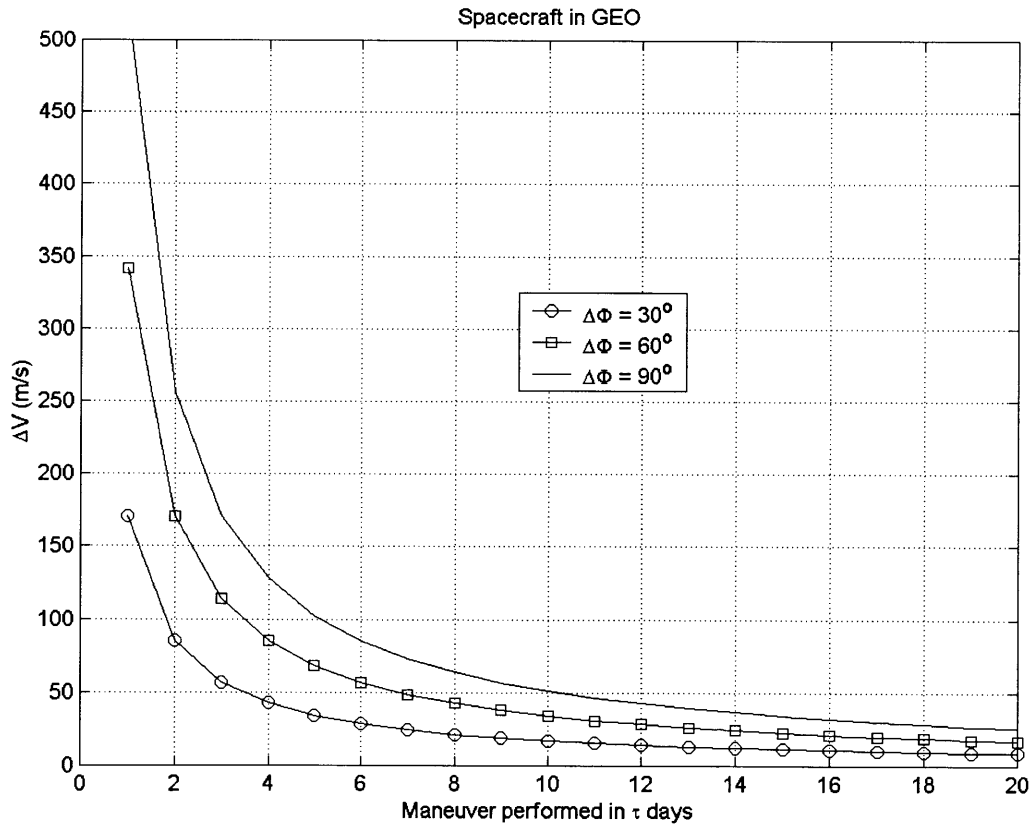
Fig. 6-11. ΔV required to perform a longitude change of $\Delta\Phi$ in τ days for a spacecraft in GEO.

174

Let ΔV_{tot} be the total velocity increment necessary to perform station keeping over the intended spacecraft design lifetime T_{Life}. If the velocity increment required to perform the phasing maneuver exceeds the fuel margin, it will reduce the actual lifetime by $\Delta T_{life\text{-}lost}$:

$$\Delta T_{life-lost} = \Gamma\left(\Delta V_{ph} - fuel\ margin\right) \times \frac{\Delta V_{ph} - fuel\ margin}{\Delta V_{tot}} \times T_{Life} \qquad (6\text{-}10)$$

$\Gamma(x)$ is a step function such that:

$$\begin{cases} \Gamma(x) = 1 \ \ for \ \ x > 0 \\ \\ 0 \ \ elsewhere \end{cases}$$

There are several ways we can translate this life reduction into a cost penalty. A simple way of doing so is to consider the spacecraft cost-per-operational day (see Chapter 4):

$$Cost_{/day} = \frac{C\left(T_{Life}\right)}{T_{Life}} \qquad (6\text{-}11)$$

The cost penalty thus incurred due to the unanticipated but necessary orbit maneuver becomes:

$$\Delta C_{penalty} = \Delta T_{life-lost} \times \left[\frac{C\left(T_{Life}\right)}{T_{Life}} \right] \qquad (6\text{-}12)$$

The customer could estimate that the spacecraft utility rate (e.g., revenues per unit time for a commercial mission) exceeds its cost-per-operational day, therefore the aggregate utility of the mission over $\Delta T_{life\text{-}lost}$ is greater than the cost penalty incurred due to the unanticipated but necessary orbital maneuver:

$$U\left[\left(T_{Life} - \Delta T_{life-lost} ; \ T_{Life}\right)\right] \geq \Delta C_{penalty} \qquad (6\text{-}13)$$

On-orbit refueling of the maneuvering spacecraft becomes cost-effective only if it can be achieved for less than $U[(t_0; \ t_1)]$. In other words, **from a customer's perspective, on-orbit refueling is worthwhile only if it costs less than the aggregate utility provided during the life extension resulting from refueling.** In the example above, we provided one simple way of

175

estimating a lower bound on the aggregate utility for a non-commercial mission. The point of this example is more to emphasize the notion of value of servicing rather than to estimate the utility aggregate provided during the life extension resulting from on-orbit refueling.

Numerical example #1: We consider the MILSTAR 2 satellite that needs to maneuver in order to cover a new theater location 90° West of its current location, in four days. The satellite cost to IOC (includes launch cost) is \$1.23b. It is designed for a 10 years lifetime. Its cost-per-operational day is:

$$Cost_{/day} = \frac{\$1.23b}{10 \times 365.25} \approx 337,000\$ \,/\, day$$

The satellite is considered to provide a service per day whose value exceeds \$337,000 (per day). The satellite is in GEO. It has a 20% fuel margin and requires approximately 52m/s for station keeping per year. The maneuver performed decreases the effective satellite lifetime by (Eq. 6-10):

$$\Delta T_{life-lost} = \frac{130 - 0.2 \times (52 \times 10)}{52 \times 10} \times 10 \times 365.25 \approx 182 \ days$$

The value of refueling the satellite at the end of its 10 years minus 182 days is worth as much as:

$$V_{refueling} \geq \Delta T_{life-lost} \times Cost_{/day} \approx \$61m$$

Similar calculations can be carried out for remote sensing or reconnaissance satellites in Low Earth Orbit: First, the maximum achievable lifetime is computed assuming no orbital maneuvers are performed and given the spacecraft propellant load. Second, the impact of an orbital maneuver (e.g., phasing maneuvers or lowering the spacecraft altitude) on the spacecraft lifetime is estimated ($\Delta T_{life-lost}$), and translated into a cost penalty. Third, assuming that the spacecraft utility rate (e.g., revenues per unit time for a commercial mission) exceeds its cost-per-operational day, the value of on-orbit refueling can be estimated using Eq. 6-12 and 6-13:

$$V_{refueling} \geq \sum_{\substack{all \\ maneuvers}} \left(\Delta T_{life-lost}\right)_i \times \frac{C(T_{Life})}{T_{Life}} \tag{6-14}$$

Numerical example #2: Let us consider in this example an astronomical observatory in Low Earth Orbit (290km × 1000km). The satellite cost to IOC (includes launch cost) is approximately $1.3b. It is designed for a 10 years lifetime. Its cost-per-operational day is:

$$Cost_{/day} = \frac{\$1.3b}{10 \times 365.25} \approx 355,900\$ / day$$

The satellite is considered to provide a service per day whose value exceeds $355,900 (per day). Orbit maintenance (atmospheric drag and J_2 effects) and station-keeping require 400m/s per year. Assume the satellite has to perform a maneuver in order to lower its perigee to 200km, then raise it back again to 290km. The maneuver consumes approximately 50m/s, or 45 days of the satellite lifetime, assuming the satellite has no fuel margin (Eq. 6-10). The value of refueling the satellite at the end of its 10 years minus 45 days is:

$$V_{refueling} \geq \Delta T_{life-lost} \times Cost_{/day} \approx \$16m$$

The value of refueling increases as the number of such orbital maneuvers increases (Eq. 6-14). The mass of the propellant required to provide a specific ΔV is given by Eq. 4-18. For a 3,000kg satellite and an I_{sp} of 300s, approximately 50kg of propellant are required to provide a ΔV of 50m/s. At 50,000$/kg to orbit, this amount of propellant can be provided to the spacecraft at roughly $2.5m. Compared with the value of refueling in our case ($16m), this result is particularly interesting and shows that on-orbit refueling is very likely to be cost-effective for high value on-orbit assets.

In the two examples discussed above, the value of refueling is found to be considerable (for the particular maneuvers considered). This results from our choice of two particularly expensive satellites (both launched on a Titan IV). It is likely however that the value of refueling for more standard satellites ($100m–$200m) would be an order of magnitude smaller. While the purpose of these examples as stated above is to emphasize the notion of value of servicing, and to illustrate one way of computing this value in the particular case of spacecraft life extension, the examples nevertheless show that refueling is likely to be cost-effective for very high-value assets. These preliminary results are very promising for the future of on-orbit refueling.

On-orbit Refueling, Time, and Risk: Decision-makers have often perceived on-orbit servicing as a significant source of technological risk. As a result, they have been reluctant to explore the

option of servicing their satellites, particularly when they were operating high-value assets. This however need not be the case: Technological risk, which we shall define in this case as the negative impacts resulting from the probability of crash/failure when attempting to dock with a host vehicle or while performing servicing, is function of the timing of the servicing activity, i.e., when does servicing occur during the lifetime of the spacecraft. According to this definition, it is **riskier** to service a newly launched spacecraft, say after one year of operations, than to service an aging spacecraft, after ten years of operations for example. Risk is minimized if servicing is performed at the end of a spacecraft lifetime when the customer can choose between end-of-life disposal or life extension through on-orbit servicing (refueling in our case). In other words, on-orbit refueling presents little risk if it is performed at the end of a spacecraft lifetime. The reader interested in a discussion on the relationship between time and risk is referred to Bernstein (1996) from which the following quote is taken:

> Risk and time are the opposite sides of the same coin, for if there was no tomorrow, there would be no risk. Time transforms risk, and the nature of risk is shaped by the time horizon: the future is the playing field.

6.6.2 Flexibility and the Value of Servicing for a Commercial Mission with Uncertain Revenues

In the previous section, we explored the concept of **value of servicing** in a simple case where the customer, a non-profit organization, sought only to minimize the cost associated with designing and operating a spacecraft, and not to maximize its profits. Two ways for computing the value of servicing in the case of spacecraft life extension were suggested. The purpose of the previous section was to illustrate the foundational idea of this new perspective on on-orbit servicing where the value of servicing, as seen from the (servicing) customer's perspective, is computed independently of any servicing architecture. For pedagogical reasons, no considerations were given to issues of flexibility, as discussed in 6.5.1 and 6.5.2. Indeed, since we did not consider any uncertainty characterizing the environment in which the spacecraft was to operate, flexibility was irrelevant: **In a world of certainty, flexibility has no value.**

In this section, we explore the value of servicing for life extension in the case of a commercial satellite with uncertain revenues. The **value of flexibility** provided by on-orbit servicing in this

case, unlike our previous calculations, can and should be accounted for in estimating the value of servicing.

The story line

Consider a commercial satellite designed for T_0 years, with an option to be serviced at T_0 in order to extend its lifetime by ΔT. We note E the cost to service the satellite (E as the exercise price of a stock option), and S the present value of the revenues generated by the satellite after T_0 (S as the stock price). The revenues S are uncertain at the time of launch; their best estimate at the time of launch ($t = 0$) is S_0. A potential customer would select on-orbit servicing for life extension only if servicing costs less than the aggregate utility provided during the life extension resulting from servicing. In other words, a customer would select to extend his/her spacecraft design lifetime if the expenses incurred for life extension and operation during ΔT are smaller than revenues generated during this same period:

$$S \geq E + C_{ops}(\Delta T) \tag{6-15}$$

$C_{ops}(\Delta T)$ is the cost to operate the satellite during ΔT. The customer's choice to exercise the option on life extension or not is captured in the decision tree of Figure 6-12.
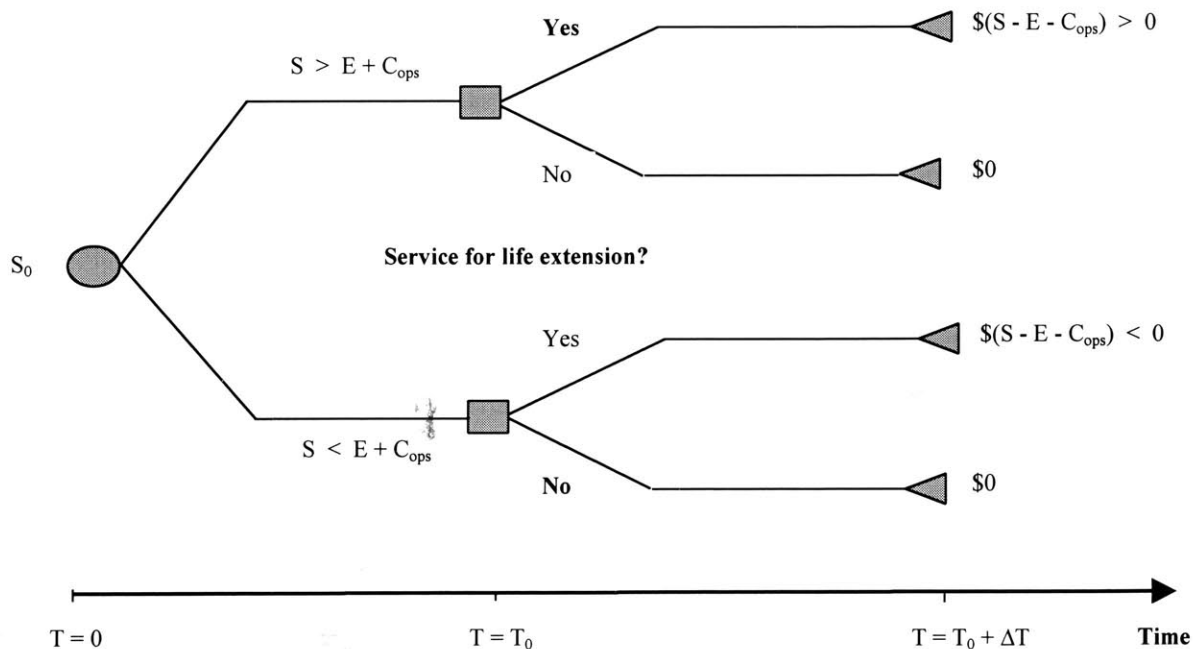
Fig. 6-12. Decision tree representing the option on life extension for a commercial satellite with uncertain revenues.

179

The situation represented in the Figure above is similar to the investment problem discussed in 6.5.2 and represented in Figure 6-7. The difference is that while the value of the project in our investment problem could take only two discrete values S^+ = \$400m and S^- = \$280m, the uncertain parameter in this example, i.e., the revenues generated after T_0, can vary within a continuous range. Therefore, an infinite number of branches shoot out of the event node. Only two however are shown on Figure 6-12 that correspond to a relevant boundary for the decision of exercising the option on life extension or not.

Let us now assume that the revenues S have a log-normal probability density function. This is a standard result in real option theory; it results from the assumption that the future value of a real asset behaves as a financial stock, therefore its rate of change can be described as a diffusion process (random walk) with volatility σ. The reader is referred to Trigeorgis (1996) for a comprehensive discussion of the diffusion process in modeling the dynamics of financial assets.

$$p(S) = \frac{1}{\sigma\sqrt{2\pi T_0}} \times \frac{S_0}{S} \times Exp\left\{ - \frac{\left[\ln\left(\frac{S}{S_0}\right) - \left(\alpha - \frac{\sigma^2}{2}\right) \times T_0\right]^2}{2\sigma^2 T_0} \right\}$$

(6-16)

σ is the volatility of the revenues after T_0, and α the expected rate of return of the revenues. We assume in the following that α is equal to the risk-free interest rate r. Equation 6-4 extended to the continuous case, provides the value of the option to service the satellite for life extension:

$$V_{DTA} = \int_0^{E+C_{ops}} 0 \times p(S) \times dS + \int_{E+C_{ops}}^{+\infty} e^{-rT_0} \times (S - E - C_{ops}) \times p(S) \times dS$$

(6-17)

Given (6-16) and (6-17), the value of the option can be written as follows:

$$V_{DTA} = S_0 \times N(d_1) - e^{-rT_0} \times (E + C_{ops}) \times N(d_2)$$

(6-18)

180

$$\text{Where} \begin{cases} N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2} \times dt \\ \\ d_1 = \dfrac{\left[\ln\left(\dfrac{S_0}{E}\right) + \left(\alpha + \dfrac{\sigma^2}{2}\right) \times T_0 \right]}{\sigma\sqrt{T_0}} \\ \\ d_2 = d_1 - \sigma\sqrt{T_0} \end{cases} \qquad N : \text{cumulative normal distribution function}$$

Equation (6-18) is identical to the Black-Scholes equation, which was a key result in the foundation of option pricing in 1973, and earned its authors the 1997 Nobel Prize in Economics.

The value of flexibility

In his 1997 Nobel Lecture, R. Merton [Merton, 97] described the relationship between uncertainty and flexibility in the following terms:

> "The future is uncertain...and in an uncertain environment, having the flexibility to decide what to do after some of that uncertainty is resolved definitely has value. Option-pricing theory provides the means for assessing that value".

Merton describes a positive correlation between uncertainty and the value of flexibility? But **how much is flexibility worth? A lot, if uncertainty is high!** Let us first explore and quantify the value of flexibility provided by on-orbit servicing in the case of life extension as a function of the volatility of the revenues σ. The value of flexibility is calculated as shown in Eq. 6-5. Figure 6-13 shows a typical result of the value of the option to service the satellite for life extension (V_{DTA}), and the value of flexibility as a function of σ.

181
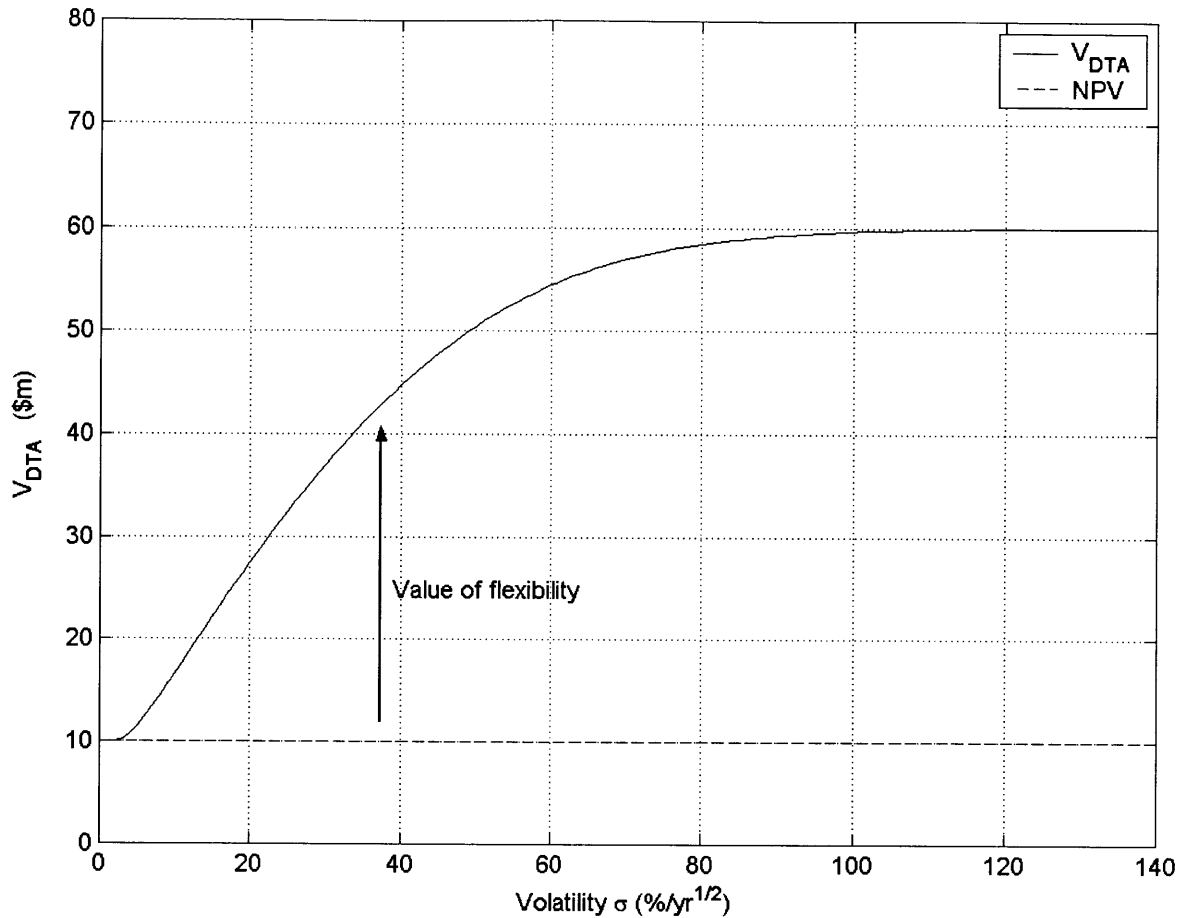
Fig. 6-13. Value of the option to service the satellite for life extension and the value of flexibility as a function of the volatility of the revenues σ (S_0 = $60m, $E + C_{ops}$ = $100m, r = 10%, T_0 = 7 years).

Figure 6-13 beautifully illustrates Merton's quote stated above: In an uncertain environment, flexibility has value. Furthermore the value of flexibility increases as the uncertainty–the volatility of the revenues in our case–increases. Figure 6-13 also shows that when there is little uncertainty on the expected revenues, option valuation (Eq. 6-18) and NPV calculation (Eq. 6-2, or the continuous version of it) yield the same result. In other words, NPV is an appropriate tool to capture the value of a project or an investment when there is little uncertainty. However, since it cannot capture the value of flexibility, it is an inadequate tool for project valuation with high uncertainty.

The discussion above has addressed the effect of volatility of the revenues on the value of the option on life extension through on-orbit servicing. In addition to the volatility, there are three

other variables that affect the value of an option (four if we consider the risk-free interest rate r). They can be easily read from Eq. 6-18. These are:

a. The present value, S, of the revenues generated by the satellite after T_0. As S increases, so does the value of the option on life extension.

b. The cost to service the satellite, E. As E increases, the value of the option to extend the life of the satellite decreases.

c. The time, T_0, when the customer decides to exercise the option to service his/her satellite for life extension or not. In financial parlance, this is called the *time to maturity* of an option. As the time to maturity increases, the value of the option increases.

In Figure 6-13, the cost to service the satellite at T_0 and to operate it for an additional ΔT years was fixed ($E + C_{ops}$ = \$100m) and the volatility was allowed to vary. This allowed a clear reading of the value of flexibility as a function of the volatility, all else being kept constant. Figure 6-14 is more complex than Figure 6-13: It represents the value of the option to service the satellite for life extension as a function of the cost to service the satellite (E) and to operate it; in the following discussion, we will call this cost ($E + C_{ops}$) the **strike price**. Several observations can be made based on Figure 6-14. First we see that the value of the option to extend the life of the satellite decreases as the strike price increases. This result is indeed intuitive and illustrates point (b) stated above. Second, we observe, as in Figure 6-13, that for a given strike price, the value of the option to extend the satellite lifetime increases with the uncertainty on the revenues during the life extension. Third, we observe that the NPV always underestimates the value of the option to service the satellite for on-orbit servicing. This results from the inability of an NPV calculation to capture the value of flexibility, as discussed previously. The value of flexibility accounts for the difference between the two valuation schemes (NPV and V_{DTA}). Fourth, we see that the maximum value of the option on life extension occurs when the strike price is zero, and is equal to the expected revenues S_0. This asymptotic behavior of V_{DTA} is readily derived from Eq. 6-18 in the following way: As $(E + C_{ops}) \rightarrow 0$, $d_1 \rightarrow \infty$, and $N(d_1) \rightarrow 1$. Therefore, the value of the option as given in Eq. 6-18 simply becomes $V_{DTA} = V_{DTA-\max} = S_0$. Finally, traditional NPV calculation establishes the existence of a boundary on the strike price (corresponding to NPV = 0) beyond which on-orbit servicing is no longer considered cost-effective. This boundary however is not valid since the value of flexibility provided by on-orbit servicing is not taken into account.
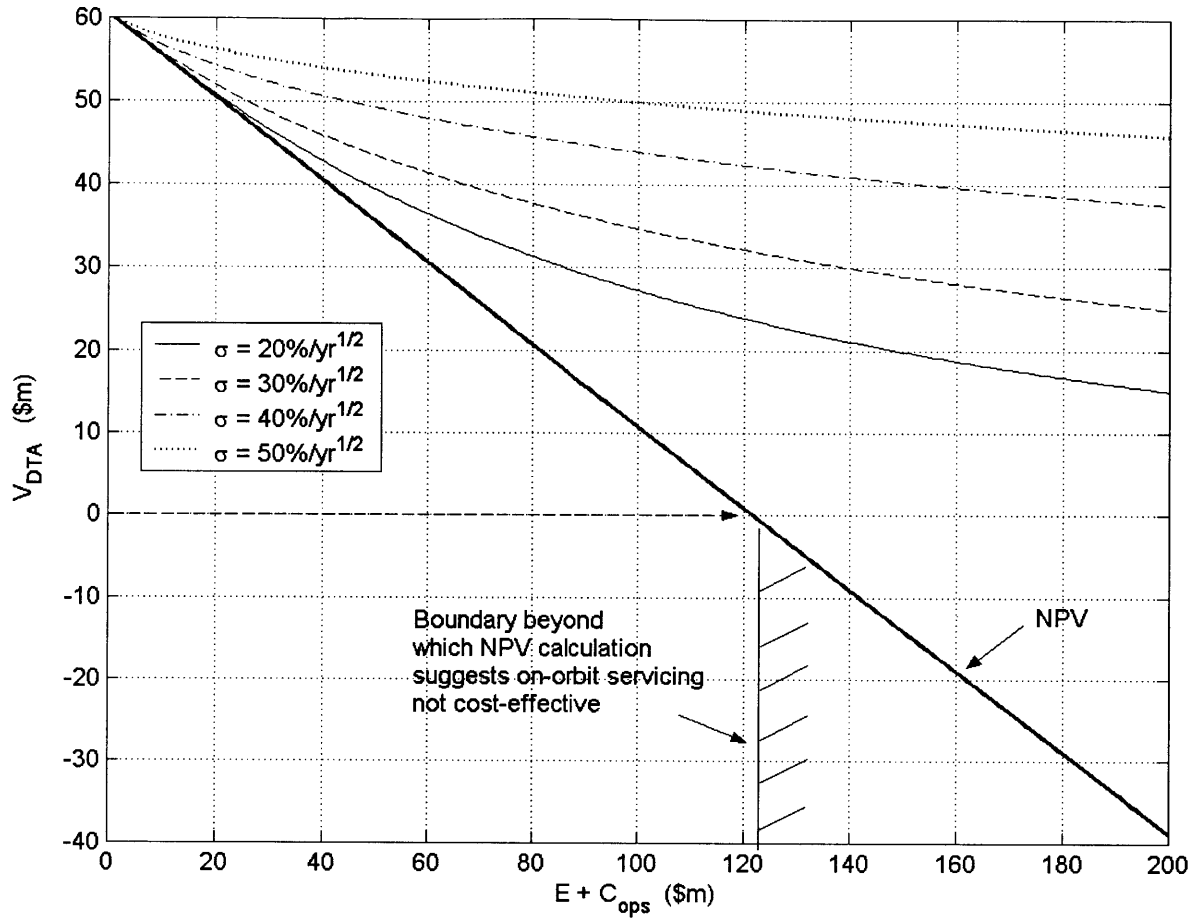
Fig. 6-14. Value of the option to service the satellite for life extension as a function of the price to service and operate it (S_0 = $60m, r = 10%, T_0 = 7 years).

The Value of Servicing

In the discussion so far, we have quantified the value of flexibility provided by on-orbit servicing, and illustrated several aspects and implications of option pricing as applied to our spacecraft life extension. However, we have not yet addressed the issue of **value of servicing** or the maximum price a customer would be willing to pay to extend the design lifetime of his/her spacecraft through on-orbit servicing. In order to do so, let us first define the **incremental value of the satellite per life extension** ΔT. This is simply equal to the expected revenues during ΔT minus the cost to design a satellite for an extra ΔT years and to operate it during this same period. Mathematically, it is written as follows:

184

$$\Delta V(\Delta T) = \int_0^\infty S \times p(S) \times dS - \left[C(T_0 + \Delta T) - C(T_0) + C_{ops}(\Delta T) \right] \qquad (6\text{-}19)$$

Recall that S is the present value of the revenues generated by the satellite during ΔT, and $C_{ops}(\Delta T)$ the cost to operate it during this same period. Equation (6-19) captures the intuition that designing a satellite for an extra ΔT years is cost-effective only if the expected revenues during this same period exceed the incremental cost for designing the satellite for an additional ΔT, i.e., when $\Delta V(\Delta T) > 0$.

We can now write the fundamental equation driving the value of servicing for spacecraft life extension in the case of a commercial system with uncertain revenues. The value of servicing in this case has been defined as the maximum price a customer would be willing to have his/her spacecraft serviced for life extension E_{max}. It is given by Eq. 6-20:

$$\underbrace{S_0 \times N(d_1) - e^{-rT_0} \times (E_{max} + C_{ops}) \times N(d_2)}_{V_{DTA}} = \underbrace{\int_0^\infty S \times p(S) \times dS - \left[C(T_0 + \Delta T) - C(T_0) + C_{ops}(\Delta T) \right]}_{\Delta V(\Delta T)} \qquad (6\text{-}20)$$

The underlying principle of Eq. 6-20 is that having the option to extend the spacecraft life should be more valuable than designing upfront for a longer design lifetime. E_{max} is the servicing price for which it is equally valuable to service the satellite than to design it upfront for an extended period. E_{max} is therefore the **maximum** servicing price a customer would be willing to pay. For a servicing price greater than E_{max}, the value of the option to extend the satellite lifetime is smaller than the value of designing the satellite upfront for a longer lifetime. This illustrates point (b) discussed above where the value of an option decreases as the strike price increases (see Fig.6-14).

For $E > E_{max} \Rightarrow V_{DTA} < \Delta V(\Delta T)$ \qquad (6-21)

We now have a way for computing the value of servicing for a commercial mission with uncertain revenues (Eq. 6-20). The parameters required to perform this calculation are recapitulated in Table 6-2.

Table 6-2. Parameters required to compute the value of servicing for life extension (Eq. 6-20).

Expected revenues during ΔT	S
Volatility of the revenues	σ
Satellite cost profile	$C(T_{Life})$
Design lifetime and life extension	T_0 and ΔT

Figure 6-15 represents a graphical solution of Eq. 6-20. The two marked points read as follows: For a $\Delta V(\Delta T) = \$48m$, the value of serving for life extension (ΔT) increases as the volatility of the expected revenues increases: It is worth $\$21m$ (minus the cost to operate the satellite during ΔT) when the volatility of the revenues σ is equal to $20\%/yr^{1/2}$, and $\$58m$ when $\sigma = 40\%/yr^{1/2}$.
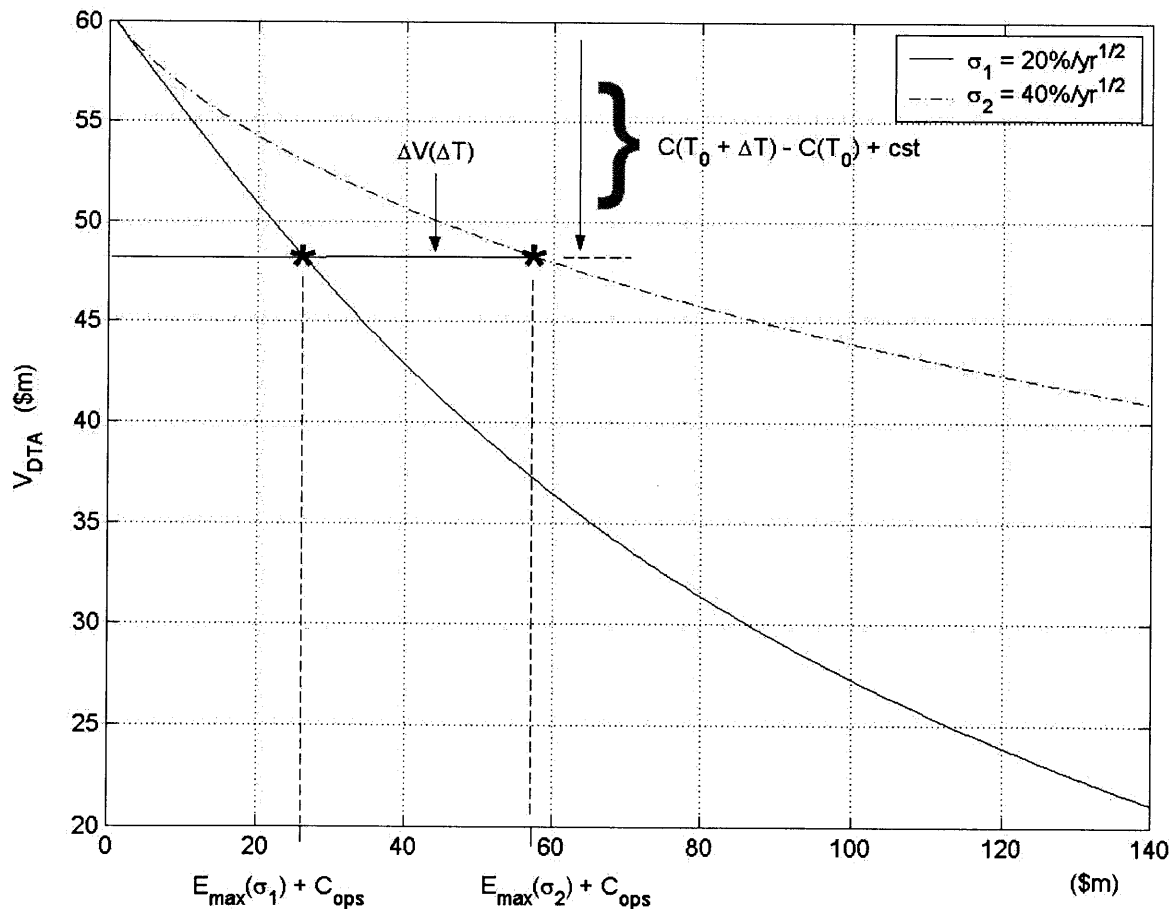
Fig. 6-15. Graphic solution of Eq. 6-20: Value of servicing as a function of the volatility of the expected revenues.

186

The main trends in the value of servicing for life extension that are captured by Eq. 6-20 and illustrated in Fig. 6-15 are the following:

1. The value of servicing increases as the volatility of the expected revenues increases (shown on Fig. 6-15).
2. The value of servicing decreases as the incremental cost to design a satellite for an extra ΔT years $C(T_0 + \Delta T) - C(T_0)$ decreases. In other words, if it doesn't cost much to design a satellite upfront for an extra ΔT years, the customer would be willing to pay very little in order to have serviced on-orbit for life extension (shown on Fig. 6-15).

6.7 Conclusions and Future Work

This paper introduced a new perspective on on-orbit servicing where the value of servicing is studied independently of its cost or any servicing architecture. Highlighting the **value** of servicing adds a new dimension to on-orbit servicing studies and shifts the focus from the traditional (servicing) provider's perspective to the (potential) customer's perspective.

The new perspective on on-orbit servicing presented here is based on three main ideas. The principal idea consists of estimating the **value of servicing** separately from its cost. The second idea lies in the observation that on-orbit **servicing provides flexibility** to space missions. And finally, contrary to what has been implicitly assumed by traditional approaches, the value of servicing is not limited to potential cost savings; instead the **value of flexibility** provided by on-orbit servicing represents an important component of the value of servicing. In other words, the third idea lies in recognizing that the value of servicing should account for the value of flexibility provided by on-orbit servicing. However, traditional discounted cash flow techniques such as the standard NPV calculation used by previous studies of on-orbit servicing cannot capture the value of flexibility. In order to circumvent this deficiency, we used Decision-Tree Analysis as a valuation tool for capturing the value of flexibility provided by on-orbit servicing.

To illustrate this new perspective, we applied it in a specific context, that of capturing the value of spacecraft lifetime extension provided by on-orbit servicing. Two ways of assessing the value of servicing were discussed. In the first case, the customer was a non-profit organization, desiring minimum cost. The value of servicing a satellite for life extension (ΔT) was derived using a cost-equivalence principle. In the second case, the customer was a for-profit organization, desiring maximum profit. The value of servicing a commercial satellite with uncertain revenues was

187

derived using a variant of the Black-Scholes equation and the incremental value of the satellite per life extension ΔT.

Regardless of the technical details or the mathematical analysis, this new perspective does not provide an argument for or against on-orbit servicing. Instead, it suggests a careful evaluation process of on-orbit servicing that focuses on the customer. Ultimately, a customer would opt for servicing if the value of servicing the spacecraft exceeds the cost of doing so, or the minimum price a provider can afford to charge for servicing. This framework identifies the value of on-orbit servicing. Future work will focus on capturing the value of flexibility and on-orbit servicing in the case of satellite upgrade or modification. This should prove particularly valuable for systems operating in a highly dynamical environment, such as an uncertain market or a fast changing technology base.

Chapter 7

Conclusions and Recommendations for Future Work

"If I lived twenty more years, and was able to work, how I
should have to modify the *Origin*, and how much the views on
all points will have to be modified! **Well it is [just] a beginning."**

C. Darwin in a letter to his friend J. Hooker on the *Origin of Species*.

7.1 Summary and Contributions

This thesis revolves around issues of flexibility in system design in general, and spacecraft design
lifetime as well as on-orbit servicing as a means for providing flexibility to space systems in
particular.

A roadmap for a comprehensive discussion of issues of flexibility in system design was proposed
that addresses the following questions: 1) What are the characteristic features of flexibility in
system design? Can one clearly and unambiguously characterize it, and disentangle it from
closely related concepts? 2) What drives the need for flexibility in system design, and what are
the attributes of an environment in which flexible designs should be sought and fielded? 3) How
can one embed flexibility in a system design? 4) What are the trade-offs associated with
designing for flexibility? What is the value of flexibility and what are the penalties (cost,
performance, risk, etc.), if any, associated with it? These are the fundamental questions around
which this thesis revolves.

The first part of this work addressed the first two questions. In order to discuss any subject matter
clearly, it is necessary to begin with a clear set of definitions. Indeed much can be gained through

careful and consistent definitions alone. Flexibility however is (was?) a word rich with ambiguity. Chapter 2 identified the various sources of ambiguity that plague discussions on flexibility (proliferation of pseudo-synonyms, timing of occurrence of "change" and attitude towards it, and the distorted perspective introduced by *Real Options* that focuses solely on the value of flexibility at the detriment of other matters). It then reviewed the concept of flexibility as discussed in various disciplines (manufacturing, finance) and extracted its characteristic features. The following definition was suggested: Flexibility (of a design) is the property of a system that allows it to respond to changes in its initial objectives and requirements–both in terms of capabilities and attributes–occurring after the system has been fielded, i.e., is in operation, in a timely and cost-effective way. Flexibility and robustness of a design were then contrasted; robustness being the property of a system that allows it to satisfy a fixed set of requirements despite changes occurring after the system has been fielded, in the environment or within the system itself. A distinction was also drawn between flexibility and universality. Flexibility of a design was also disentangled from issues of flexibility in the design process, the latter including activities, methods, and tools devised to mitigate the risks–cost, schedule, and performance–resulting from requirement changes occurring during the design process, i.e., before a system is fielded.

In Chapter 3, it was argued that flexibility should be sought: 1) when the uncertainty in a system's environment is such that there is a need to mitigate market risks, in the case of a commercial venture, and reduce a design's exposure to uncertainty in its environment, 2) when the system's technology base evolves on a time scale considerably shorter than the system's design lifetime, thus requiring a solution for mitigating risks associated with technology obsolescence. In other words, two fundamental consequences of the property of flexibility (of a design) were identified: flexibility reduces a design's exposure to uncertainty, and provides a solution for mitigating market risks as well as risks associated with technology obsolescence. A metric was introduced that quantifies the disparity between components' life cycle (or Time to Obsolescence of a component) and the system's design lifetime, thus driving the need for flexibility.

One way through which flexibility manifests its criticality to systems architects is in the specification of the system design lifetime requirement. The second part of this work addressed issues of design lifetime, and ways to provide and value flexibility in the particular case of space systems.

Chapter 4 explored the impacts of the design lifetime requirement on spacecraft mass and cost to IOC. First, it was shown that design lifetime is a key requirement in sizing various spacecraft subsystems. Second, spacecraft cost profiles as a function of the design lifetime were established and a cost per operational day metric was introduced. It was found for instance that a cost penalty of 30% to 40% is incurred when designing a spacecraft for fifteen years instead of three years, all else being equal, and that the cost per operational day decreases monotonically as a function of the spacecraft design lifetime. In the absence of other metrics, this result justifies pushing the boundary for increasingly longer spacecraft design lifetimes and suggests that a customer is always better off asking the contractor to provide the maximum design lifetime technically achievable. The following chapter proves this intuition to be wrong.

Chapter 5 argued for an augmented perspective on system architecture (diachronic) that complements the traditional views on system architecture (synchronic). It was suggested for instance that the design lifetime is a fundamental component of a system's architecture although one cannot see it or touch it. Consequently, cost, utility, and value per unit time metrics were introduced and explored in order to identify optimal design lifetimes for complex systems in general, and space systems in particular. It was found that an optimal design lifetime for a satellite exists, even in the case of constant expected revenues per day over the system's lifetime, and that it changes substantially with the expected Time to Obsolescence of the system and the volatility of the market the system is serving in the case of a commercial venture. The analysis thus proved that it is essential for a system architect to match the design lifetime with the dynamical characteristics of the environment the system is/will be operating in. It was also shown that as the uncertainty in the dynamical characteristics of the environment the system is operating in increases, the value of having the option to upgrade, modify, or extend the lifetime of a system at a later point in time increases depending on how events unfold.

On-orbit servicing provides a way to physically access, upgrade, and modify a spacecraft. In other words, on-orbit servicing provides flexibility to space systems. Chapter 6 developed a new perspective on on-orbit servicing that builds on the concepts and the results discussed in all the previous chapters. This new perspective is based on three main ideas: The principal idea consists of estimating the value of servicing, as seen from the customer's perspective, independently from its cost or specific implementation. The second ideas lies in the observation that on-orbit servicing provides flexibility to space systems. The third ideas recognizes that the value of servicing, contrary to what has been implicitly assumed by the traditional approaches, is not limited to cost savings. Instead, it is shown that the value of flexibility provided by on-orbit

servicing is an important component of the value of servicing. A valuation tool that leverages the advantages of Decision-Tree Analysis and Real Options is developed that captures the value of this flexibility. To illustrate this new perspective, it was applied to the specific case of capturing the value of spacecraft lifetime extension provided by on-orbit servicing. Two ways of assessing the value of servicing were discussed. In the first case, the customer was a non-profit organization, desiring minimum cost. The value of servicing a satellite for life extension (ΔT) was derived using a cost-equivalence principle. In the second case, the customer was a for-profit organization, desiring maximum profit. The value of servicing a commercial satellite with uncertain revenues was derived using a variant of the Black-Scholes equation and the incremental value of the satellite per life extension ΔT. Regardless of the mathematical details involved, and while the results are very encouraging for the future of on-orbit servicing, this new perspective does not provide an argument for or against on-orbit servicing. Instead, it suggests a careful evaluation process of on-orbit servicing that focuses on the customer. Ultimately, a customer would opt for servicing if the value of servicing the spacecraft exceeds the cost of doing so, or the minimum price a provider can afford to charge for servicing.

Asides from the particular contributions per chapter mentioned above, there are two conceptual contributions that are not discussed at any one point in this thesis; instead they are pervasive throughout the whole work. These are discussed below.

A fundamental conceptual contribution of this thesis is in the introduction of time considerations into system architecture. Using the terminology introduced in the Introduction and borrowed from linguistics, this is equivalent to introducing a **diachronic perspective on system architecture,** when traditionally the synchronic approach (the "snapshot" approach) has prevailed. System architecting has been traditionally viewed as a matching between two (vector) quantities, resources and system's performance. One approach fixes the amount of available resources and strives to maximize the system's performance; the other approach constrains the system's performance and attempts to minimize the resources necessary to achieve the target performance [de Weck, 01]. The first approach operates with–and attempts to maximize–a performance per unit cost metric; the second approach seeks to minimize a cost per function (or performance) metric. This thesis proposed to view in a system architecture the flow of service (or utility) that the system will provide over its design lifetime. Consequently, cost, utility, and value per unit time metrics were introduced. It therefore suggests that we augment our understanding of system architecture by considering the system's design lifetime, as well as other time characteristics associated with a design, as fundamental components of system architecture although one cannot

192

see them or touch them. One direct consequence of this perspective is that space operations for example are part of a space system architecture and should be addressed upfront in the design process.

A second conceptual contribution of this thesis is in recognizing the fundamental relationships between Time, Uncertainty, and Flexibility; the three faces of a same coin. Time and uncertainty are intrinsically related, for if there were no tomorrow, there would be no uncertainty. Time transforms uncertainty, which in turn is shaped by the time horizon [Bernstein, 96]. Flexibility on the other hand reduces the exposure to uncertainty thus allows a system to weather the heavy hand of Time.

7.2 Recommendations for Future Work

Somebody once said, "a thesis is never completed, it is always abandoned": this is where I chose to–and my committee agreed that I can–*abandon* my thesis. This work however has unearthed more interesting questions than it has answered. In the following, some of the thought-provoking questions I have stumbled upon but did not address are presented and grouped into three categories: Conceptual/Theoretical, Practical/Data Collection and Analysis, and a third category related to On-Orbit Servicing that includes both a conceptual and a practical slant to it. I am confident that whoever chooses to pursue any one topic will find it interesting, challenging, and useful to the technical community at large.

7.2.1 Conceptual/Theoretical

1. A Framework for Identifying Flexibility-Enabling Practices

Assume that a system can operate in $\underline{n}$ different modes; a boiler for instance that can be fired using oil or gas, or a space system that can perform multiple roles such as GMTI, SigInt, and Geo-location:

$$\underline{m} = [m_0, m_1, ..., m_n]^T \tag{7-1}$$

We can construct a **flexibility cost-penalty vector** as follows: Each element in the vector is the ratio of the cost of a system that operates in all modes $\underline{n}$ and one that operates only in mode $m_i{}^1$.

---

[1] Assuming same performance in mode m_i.

$$p_{flx} = \left[\frac{C(\underline{n})}{C(m_1)}, \frac{C(\underline{n})}{C(m_2)}, ..., \frac{C(\underline{n})}{C(m_n)} \right]^T \qquad (7\text{-}2)$$

This vector represents the cost penalty for embedding one type of flexibility in the design. For instance, in the case of industrial boilers that can be fired by either gas, oil, or both, assuming we have the following purchase prices:

Table 7-1. Purchase price of three different boilers

Boiler Type	Cost (FY00$)
Gas	60,000
Oil	65,000
Dual-fuel	75,000

The flexibility cost-penalty vector is:

$$p_{flx} = \left[\frac{75}{60}, \frac{75}{65} \right]^T = [1.25, \ 1.15]^T \qquad (7\text{-}3)$$

This vector is important when evaluating the usefulness of embedding several modes of operations in a system design. For instance, let us consider a system that can operate in mode m_1 and mode m_2. Acquiring such a system would make sense if, all else being equal:

$$\chi_c = \frac{C(m_1 + m_2)}{C(m_1) + C(m_2)} \leq 1 \qquad (7\text{-}4)$$

Otherwise it is less expensive to acquire separately designs that operate in single modes m_1 or m_2. We call χ_c the **cost compressibility of a design**. A design practice that enables or provides flexibility to a design reduces χ_c (e.g., modularity).

Continuing along similar lines, since flexibility was characterized by the ease of change of a design to satisfy different requirements at different points in time, we can define a **switching cost matrix** to reflect the cost-effectiveness of a system reconfiguration to operate in a different mode: Given $\underline{n}$ modes of operation for a design, the elements $c_{i,j}$ of the matrix represents the cost to reconfigure the system from operating in mode i to mode j. By definition, $c_{i,i} = 0$, no switching is involved in remaining in the same mode of operation:

194

$$C = \begin{bmatrix} 0 & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & 0 & \cdots & c_{2,n} \\ \cdots & \cdots & 0 & \cdots \\ c_{n,1} & c_{n,2} & \cdots & 0 \end{bmatrix} \qquad (7\text{-}5)$$

A similar matrix can be constructed that represents the time required to reconfigure the system from operating in mode i to mode j:

$$T = \begin{bmatrix} 0 & \tau_{1,2} & \cdots & \tau_{1,n} \\ \tau_{2,1} & 0 & \cdots & \tau_{2,n} \\ \cdots & \cdots & 0 & \cdots \\ \tau_{n,1} & \tau_{n,2} & \cdots & 0 \end{bmatrix} \qquad (7\text{-}6)$$

These two matrices are good indicators of system's flexibility in the case of multiple modes of operations, that is when the different requirements to be satisfied by the system at different points in time are known a priori: The smaller the elements in the matrices, the more flexible the system is for a given change (ease of change). In the case of the multiple role TechSat21 mission for example, the elements in the above matrices would be the cost—in terms of propellant burn, ground operators assistance, etc—and time required to reconfigure the cluster from a radar mode to a geo-location mode.

Flexibility-enabling practices (e.g., modularity, scalability, platform design) can now be explored in the light of the above discussion, in particular through their impact on the cost compressibility of a design (χ_c) and the switching matrices (C, T). The impact of such practices is conceptually represented on Figure 7-1.
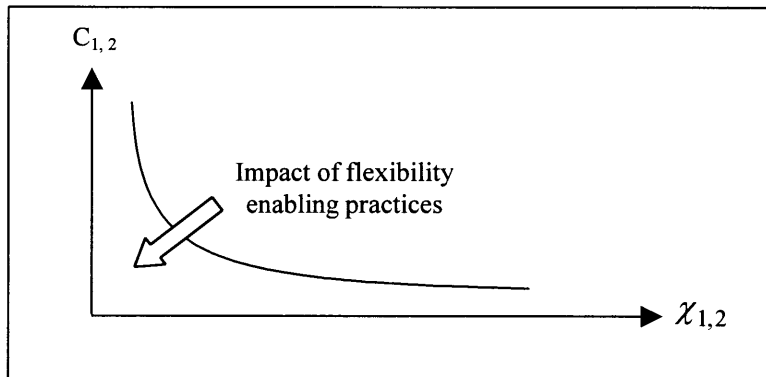
Fig. 7-1 Impact of flexibility-enabling practices on design cost compressibility and switching matrices

195

2. Flexibility and Universality versus Complexity of a Design

In Chapter 2, a distinction was drawn between flexibility and universality of a design. Software for instance that can be used in a variety of situations without change or modification, is considered "universal" not flexible. *Flexible* software [Parnas, 78] is one that can be easily changed–extended, contracted, or else–in order to be used in a variety of ways. Similarly, spacecraft that carry multiple instruments and perform multiple missions simultaneously are NOT considered flexible according to the definition of flexibility provided in this thesis. Likewise, a design is considered flexible if it is easily changeable to be used in a variety of ways. The time and cost required to implement the changes are two indicators of the "ease of change" of a design and reflect its flexibility. An interesting area to explore is the trade-off between flexibility or universality of a design and its complexity. A conceptual such relationship is represented on Figure 7-2.
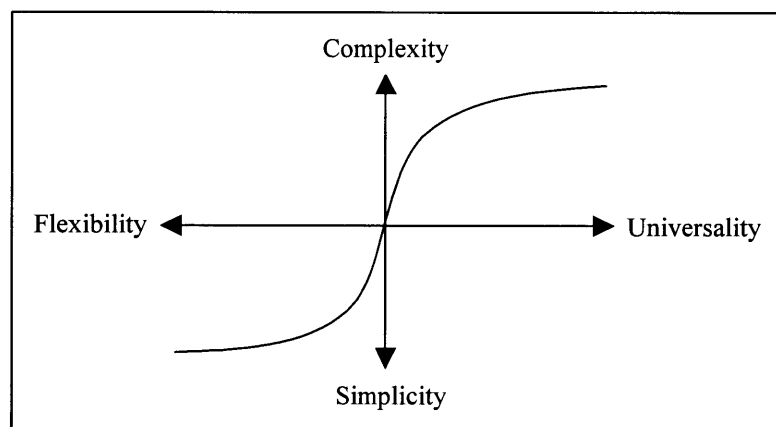
Fig. 7-2 Trade-off between flexibility or universality of a design and its complexity.

3. Flexibility and Distributed Architectures versus Monolithic Designs

To date, the world of automation is mostly single agent, *individualistic*. It is however fair to ask, "when is it better to go alone, and when is it better to have teammates?" This question applies not only to human endeavors, but also to robotics, space systems and interplanetary exploration. There are many reasons why a distributed collaborative architecture is better suited for certain applications, in particular for space applications[2]. In the case of TechSat21 for example, it was

---

[2] Constructing tools from a collection of individuals is not a novel endeavor. A chain is a collection of links, a rake a collection of tines, and a broom a collection of bristles. Sweeping the sidewalk would certainly be difficult with a single or even a few bristles [Kube and Zhang, 1994].

shown in Chapter 2 that modifying the cluster geometry allows the user to modify the revisit time, and to switch between a radar mode and geo-location mode. These features characteristic of flexibility are made possible by the distributed architecture of the system and are not possible with a monolithic design. Two research directions can be explored based on this observation that have the following objectives: First to contribute to principled synthesis of group behavior in heterogeneous multi-agent systems, in the particular case of space systems and interplanetary exploration. Second to investigate the effect of distributedness on system's flexibility. It can be hypothesized that designing for collective behavior enhances both system's robustness and flexibility.

7.2.2 Practical/Data Collection and Analysis

1. Requirement Specifications and Design Choices in the Case of the B-52 versus B-58

In order to illustrate the relationship between a product's lifetime, the initial circumstances from which the system's requirements were derived and the various environments in which it can operate, the Boeing B-52 Stratofortress was presented in Chapter 2 as an example of a flexible design (although its flexibility is perhaps accidental), and contrasted with the Convair B-58 Hustler.

The B-52 is a long-range, heavy bomber that can perform a variety of missions. It is capable of dropping or launching the widest array of weapons in the U.S. inventory. It is referred to as the bomber that "is not getting older, just getting better" because it was capable of accommodating numerous improvements over the years. The B-52 first entered service in 1955 with the Strategic Air Command. Current engineering analysis shows the B-52 lifespan can be extended beyond the year 2045.

The Convair B-58 Hustler on the other hand was the first supersonic bomber to enter service with the USAF in March 1960. Despite its high performance and sophisticated equipment, the service of the B-58 was brief; the aircraft flew for only a decade before being consigned to storage. Part of the reason for this rather short service was due to the aircraft's rather high accident rate. Another factor was the intercontinental ballistic missile, which entered service at the same time as the B-58 and removed its primary mission. Of course the same was true of the B-52 but it proved flexible enough to find widespread use in other mission areas. Aside from the technical problems that plagued the B-58, the aircraft in some sense lacked the flexibility of the B-52 to adapt to new missions and roles in new environments.

It would be interesting to probe the original requirements of both the B-52 and B-58, and to identify the particular design choices that rendered on one hand the B-52 a flexible design to remain in operation for almost a century, and on the other hand the B-58 a short-lived inflexible design. The study should investigate for example the impact of the requirement to fly at supersonic speeds for the B-58 on the wing design and the airframe, and how this choice, later during the operational life of the B-58, prevented it from accommodating different weapons and performing other missions than the one the it was initially designed for.

2. Story of Failure: Iridium, Market dynamics, and the Lack of Flexibility

Iridium is one of the biggest technological gamble and dramatic failure of the commercial space systems. The 66-satellite telephony system entered service in 1998–10 years after it was conceived–and filed for bankruptcy in 1999 after sinking over $3 billions. Lessons from its failure although not yet fully extracted, would illustrate, among other things, the need for flexibility in high-value assets as discussed in this work. What went wrong? The target market of Iridium changed between the time the business plan was laid out and when the system became operational: The cellular market took off, as did the market for data. By targeting only the market of business travelers, Iridium set itself up against the cellular players. So by the time Iridium entered operation, the cellular-phone technology had overtaken it. The market analysis performed by the system designers identified and explored a steady state or equilibrium configuration of the market; it failed to identify the dynamic nature of its market and did not embed in the system the ability to track a dynamic market and changing customers' needs. Furthermore, its handset was seen to be heavy and outdated, its voice quality not very good, and most importantly, its cost prohibitive compared to other services. In addition, the inability of the Iridium to transfer data proved a serious shortcoming of the system in the age of the Internet.

The analysis of the failure of Iridium should prove to be a rich topic, yet little explored (lessons from the failure have not yet been extracted, or at least not published). It would be interesting to obtain Iridium's initial market forecast and correlate its findings with the architectural choices that sealed the fate of the system. Were there other design choices that would have permitted the system to track new requirements arising from a new market environment and prevent its untimely obsolescence?

7.2.3 On-Orbit Servicing Studies

Two distinct thrusts can be pursued in relation to on-orbit servicing: a conceptual approach and a practical approach. It is suggested that these two approaches be jointly performed in order to produce a noticeable and useful impact.

1. Conceptual: Value of On-Orbit Servicing in the Case of Upgrade and Modification

In order to illustrate the new perspective on on-orbit servicing presented in Chapter 6, a specific case was considered of capturing the value of spacecraft lifetime extension provided by on-orbit servicing. Two ways of assessing the value of servicing were discussed. In the first case, the customer was a non-profit organization, desiring minimum cost. The value of servicing a satellite for life extension (ΔT) was derived using a cost-equivalence principle. In the second case, the customer was a for-profit organization, desiring maximum profit. The value of servicing a commercial satellite with uncertain revenues was derived using a variant of the Black-Scholes equation and the incremental value of the satellite per life extension ΔT.

Future work should focus on capturing the value of flexibility and on-orbit servicing in the case of satellite upgrade or modification. This should prove particularly valuable for systems operating in highly dynamical environments, such as an uncertain market or fast changing technology base. The analysis performed in Chapter 5 on optimal spacecraft lifetime under market uncertainty and Time to Obsolescence should prove crucial for this task.

2. Potential Customer Survey

In order for on-orbit servicing to become a common reality, it is important to have real potential customers involved in the valuation process. It is therefore preferable that the work on on-orbit servicing discussed here not remains confined to an academic environment. It would be useful to identify potential customers of on-orbit servicing, to survey them and (help them) capture their rationale in identifying the value of servicing.

3. Software Maintenance and Upload

Three different generic ways can be thought of for modifying or providing new functionalities to a system: 1) physically accessing the system and modifying or adding new modules/functions, 2) remotely evolving the hardware in order to generate new functionalities (evolvable hardware), 3) remotely upgrading or modifying the system's on-board software in order to generate new or

enhanced capabilities. While the majority of complex engineering systems take advantage of logistics and maintenance support; aircraft operational lifetime and capabilities for example are extended through routine maintenance and payload upgrade because they are physically accessible, satellites remain one of the rare systems without maintenance, repair, and upgrade infrastructure. In this work, only physically accessing a satellite was considered to provide on-orbit servicing. It is certainly worth exploring the two other options listed above, remotely evolving the on-board hardware and uploading new software, as practices that provide flexibility to space systems. In particular, upgrading or modifying a satellite's on-board software can be rightfully subsumed under on-orbit servicing, although of a different kind than the one discussed in Chapter 6. The Center for Integrated Space Microsystems (CISM) at JPL published in November 2000 the first paper on Guarded Software Upgrade (GSU) architecture design for deep space missions [Alkalai *et* al., 00]. This should prove a very rich area to explore in the case of Earth-orbiting, scientific, military, and commercial missions (requirements and technical implementations, limits and boundaries of software-dependant capabilities, value and trade-offs, etc).

7.3 Closing Remarks

Time is the fundamental object that links all the parts of this thesis. In Chapter 1, a parallel was drawn between the ephemeral nature of human life and the transiency of human handiwork.

> The ephemeral nature of human life has been a major theme for philosophers, theologians, poets, and others, ever since the dawn of history. A myriad of human behaviors and artifacts stem from, or find the original impetus for their existence in an individual or collectivity's relationship with *time*. Just like "we are [physiologically] the children of gravity, we cannot see it or touch it, but it has guided the evolutionary destiny of every species, and has dictated the size and shape or our organs and limbs[3]", so are many of our psychological dispositions, behaviors, and constructs the children of our relationship with *time* (we cannot see it or touch it...) and the recognition of the transiency of human life.

> Less profound but no less thought-provoking, is the transiency of human handiwork. Of all the structures and artifacts of antiquity, only an infinitesimal remnant survives today [Terborgh, 49]. Examples abound as well in more recent periods of industries,

---

[3] D. Newman. "Human Spaceflight from MIR to Mars". AIAA-SF June 2000 Dinner Meeting. Sunnyvale.

equipments, and products that exhibit an **ephemeral relationship with Time**, and stand as modern reminders of the transiency of such artifacts. Typically a product progresses through a life cycle characterized by periods of growth, maturity, and decline, then it dies-out because of physical, functional, or economical degradation or inadequacy. At Cape Canaveral for instance, lie the remnants of the race to the Moon: concrete launch pads, bunkers, and steel gantries in ruins from the Mercury, Gemini, and Apollo missions. Similarly, outside Tucson, in the Arizona desert, one finds the Aerospace Maintenance and Regeneration Center (AMARC), better known as the aircraft graveyard where over four thousand aircraft lie moldering in the sun. These modern ruins, familiar technological objects, stand as reminders that nothing is permanent. Through physical or functional degradation, or loss of economic usefulness, the hand of time lies heavy on the work of humans.

Fundamentally, it was shown that both living organisms and complex engineering systems strive in an ever-changing, competitively aggressive environment. Organisms or systems that are better equipped to adapt to changing environments, i.e., flexible systems or organisms, live longer or outlive more *rigid* systems or organisms. Instead of passively observing the effects of time on a fielded system (physical, functional, or economic degradation), the underlying philosophy advocated in this thesis has been to actively manage a system's relationship with time after it has been fielded. This was achieved by bringing issues of flexibility upfront in the design process, and striving to embed flexibility in the design. While the study of uncertainty in a particular environment is useful in and of itself, it is nevertheless a descriptive task of little use to decision makers. Shifting the emphasis from uncertainty to flexibility presents several advantages in this respect. While the study of flexibility still involves an analysis of the uncertainty in a system's environment, it nevertheless seeks to provide options for dealing with different events after some of the uncertainty is resolved. It is therefore a more active way of interacting (or designing a system to interact) with an environment, and should prove more appealing to decision-makers than a mere study of uncertainty.

Epilogue

"Demeurent en tout cas bien des zones d'ombres que le temps n'a fait qu'épaissir...On peut imaginer qu'à l'issue de sa conversation avec Nader...[Tanios] hésitait. On pourrait même énumérer les raisons qui avaient pu l'inciter à partir et celles au contraire qui auraient dû le retenir. A quoi bon? Ce n'est pas ainsi que se prend la décision de partir. On n'évalue pas, on n'aligne pas avantages et inconvénients...

A ce point de mes tatonnements, j'avais un peu oublié le trouble de Tanios devant mon propre trouble...J'en étais même arrivé a me dire qu'il y avait peut-être après tout quelque sortilège attaché au rocher de Tanios. Lorsqu'il était revenu s'y assoir, ce n'était pas dans le but de réfléchir, me dis-je, ni de peser le pour et le contre. C'est de tout autre chose qu'il ressentait le besoin. La méditation? La contemplation? Plus que cela, la décantation de l'âme."

Amin Maalouf in "Le Rocher De Tanios", 1993.

References

Chapter 1

[Amram and Kulatilaka, 99]   M. Amram, N. Kulatilaka. *Real Options: Managing Strategic Investments in an Uncertain World*. Harvard Business School Press, 1999.

[Bernstein, 96]   P. Bernstein. *Against the Gods: The Remarkable Story of Risk*. Wiley and Sons, 1996.

[Darwin, *Origin of Species*]   C. Darwin. *On the Origin of Species*. Introduced by R. Leakey. Hill and Wang, NY, 1979.

[Fine, 98]   C. H. Fine. *Clockspeed: Winning Industry Control in the Age of Temporary Advantage*. Perseus Books, 1998.

[Finch, 00]   G. Finch. Linguistics Terms and Concepts. St Martin's Press, NY, 2000.

[Fox and Yeh, 99]   M. A. Fox, B. P. Yeh. "Intelligent Kinetic Systems in Architecture", *in* Managing Interactions in Smart Environments (MANSE). Springer-Verlag, London, 1999.

[Highsmith III, 99]   J. A. Highsmith III. *Adaptive Software Development: A Collaborative Approach To Managing Complex Systems*. Dorset House Publishing, 1999.

[McKinon, 78]   R. McKinon. "The Concept of Flexibility and Urban Systems". Center of Urban Studies, University of Illinois, 1978.

[Parnas, 79]   D. Parnas. "Designing software for Ease of Extension and Contraction". IEEE Transactions on Software Engineering, Volume SE-5, Number 2, March 1979. pp. 128-138.

[Raouf and Ben-Daya, 95]   A. Raouf, M. Ben-Daya (Editors). *Flexible Manufacturing Systems: Recent Developments*. Elsevier, 1995.

[de Saussure, 66]   F. de Saussure. *Course in General Linguistics*. McGraw-Hill, 1966 (first published in 1915).

[Terborgh, 49]   G. Terborgh. *Dynamic Equipment Policy*. McGraw-Hill, 1949.

[Trigeorgis, 96]   L. Trigeorgis. *Real Options: Managerial Flexibility and Strategy in Resource Allocation*. MIT Press, 1996.

[de Weck, 01]   O. de Weck. "Multivariable Isoperformance Methodology for Precision Opto-Mechanical Systems". PhD thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 2001.

Chapter 2

[Ackerman, 93]   J. Ackerman. *Robust Control: Systems with Uncertain Physical Parameters*. Springer-Verlag, 1993.

[Amram and Kulatilaka, 99] M. Amram, N. Kulatilaka. *Real Options: Managing Strategic Investments in an Uncertain World*. Harvard Business School Press, 1999.

[Blanchard, 98]   B. S. Blanchard. *System Engineering and Management*. Wiley-Interscience, 1998.

[Chen and Lewis, 99]   W. Chen, K. Lewis. "Robust Design Approach for Achieving Flexibility in Multidisciplinary Design". AIAA Journal, Vol. 37, No. 8, August 1999. pp. 982-989.

[Chen and Yuan, 99]   W. Chen, C. Yuan. "A Probabilistic-Based Design Model for Achieving Flexibility in Design". Journal of Mechanical Design, Vol.121, March 1999. pp. 77– 83.

[Clausing, 94]   D. Clausing. *Total Quality Development: A Step-by-Step Guide to World-Class Concurrent Engineering*. ASME Press, New York, 1994.

[Doyle et al., 92]   J. C. Doyle, B. Francis, A. Tannenbaum. *Feedback Control Theory*. Maxwell/Macmillan, 1992.

[Dahleh and Diaz-Bobillo, 95] M. A. Dahleh, I. J. Diaz-Bobillo. *Control of Uncertain Systems: A Linear Programming Approach*. Prentice-Hall, Englewood Cliffs, NJ, 1995.

[Faulkner, 96]   T. Faulkner. "Applying Options Thinking to R&D Valuation". Research-Technology Management, May-June 1996, 39(3).

[Fox and Yeh, 99]   M. A. Fox, B. P. Yeh. "Intelligent Kinetic Systems in Architecture", *in* Managing Interactions in Smart Environments (MANSE). Springer-Verlag, London, 1999.

[Francis, 87]   B. Francis. *A Course in H_∞ Control Theory*. Springer-Verlag, Berlin, 1987.

[Fricke *et* al., 00]   E. Fricke, A. Schulz, S. Wenzel, H. Negele. "Design for Changeability of Integrated Systems within a Hyper-Competitive Environment," in INCOSE Colorado 2000 Conference, Denver, March 2000.

[GAO-01-288, 01]   General Accounting Office. "Best Practices: Better Matching of Needs and Resources will Lead to Better Weapon System Outcomes". GAO-01-288, March 2001.

[Highsmith III, 99]   J. A. Highsmith III. *Adaptive Software Development: A Collaborative Approach To Managing Complex Systems*. Dorset House Publishing, 1999.

[Hazelrigg, 96]   G. A. Hazelrigg. *Systems Engineering: An Approach to Information-Based Design*. Prentice-Hall, 1996.

[Hazelrigg, 99]   G. A. Hazelrigg. "An Axiomatic Framework for Engineering Design". Journal of Mechanical Design, Vol. 121, September 1999.

[Holmes, 99]   C. Holmes. "Effective Platform Designs for Medium Lift Helicopters". S.M. thesis. Systems Design and Management, Massachusetts Institute of Technology, 1999.

[IEEE, 98]   IEEE Guide for Developing System Requirements Specifications. 1998 Edition.

[JPL D-11509, 1996]   Mars Global Surveyor: Spacecraft Requirements. JPL D-11509, 1996.

[Lewis and Mistree, 99]   K. Lewis, F. Mistree. "Modeling Interactions in Multidisciplinary Design: A Game Theoretic Approach". AIAA Journal, Vol. 35, No. 8, June 1999, pp. 1387 – 1392.

[Martin and Stallard, 99]   M. Martin and M. Stallard. "Distributed Satellite Missions and Technologies: The TechSat21 Program". AIAA-99-4479.

[McKinon, 78]   R. McKinon. "The Concept of Flexibility and Urban Systems". Center of Urban Studies, University of Illinois, 1978.

[Mohandas and Sandgren, 89]   S. U. Mohandas, E. Sandgren. "Multiobjective Optimization Dealing With Uncertainty". Advances in Design Automation, Vol. 19-2: 1989. pp. 241 – 248.

[NASA, SP-6105]   NASA Systems Engineering Handbook, SP-6105, June 1995.

[Olson, 98]   J. Oleson. *Pathways to Agility*. Wiley, 1998.

[Parnas, 79]   D. Parnas. "Designing software for Ease of Extension and Contraction". IEEE Transactions on Software Engineering, Volume SE-5, Number 2, March 1979. pp. 128-138.

[Phadke, 89]   M. S. Phadke. *Quality Engineering Using Robust Design*. Prentice-Hall, 1989.

[Piore, 89]   M. Piore. "Corporate Reform in American Manufacturing and the Challenge to Economic Reform". Mimeo, Massachusetts Institute of Technology, 1989

[Raouf and Ben-Daya, 95]   A. Raouf, M. Ben-Daya (Editors). *Flexible Manufacturing Systems: Recent Developments*. Elsevier, 1995.

[Rappaport, 69]   A. Rappaport *Two-Person Game Theory*. University of Michigan Press, 1969.

[Sanchez-Pena and Sznaier, 98] R. Sanchez-Pena, M. Sznaier. *Robust Systems: Theory and Applications*. Wiley, 1998.

[Shaw, 99]   G. B Shaw. "The Generalized Information Network Analysis Methodology for Distributed Satellite Systems". PhD thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, February 1999.

[Stefani et al., 94]   R. T. Stefani, C. J. Savant, B. Shahian, G. H. Hostetter. *Design of Feedback Control Systems*. Saunders College Publishing, 1994.

[Suarez and Cusumano, 91] F. Suarez, M. Cusumano, C. Fine. "Flexibility and Performance: A Literature Critique and Strategic Framework". Sloan School WP 3298-91-BPS, Massachusetts Institute of Technology, 1991

[Suh, 90]   N. P. Suh. *Principles of Design*. Oxford University Press, 1990.

[Suh, 99]   N. P. Suh. *Axiomatic Design: Advances and Applications*. Oxford University Press. (to be published). Manuscript format, 1999.

[SWELT: RM0.3] Requirements Management Guidebook. SWELT: RM0.3 9/30/96.

[Tagushi, 87]   G. Taguchi. *Systems of Experimental Design* (English translation, edited by D. Clausing). UNIPUB/Kraus International Publications, New York, 1987.

[Taylor, 91]   T. Taylor. "Evaluating and Selecting Manufacturing Flexibility". S.M. thesis. Department of Mechanical Engineering, Massachusetts Institute of Technology, 1991.

[Thurnton, 91]   D. L. Thurston. "A Formal Method for Subjective Design Evaluation with Multiple Attributes". Research in Engineering Design, Vol. 3: 1991. pp. 105 – 122.

[Triantis and Hodder, 90] A. Triantis, J. Hodder. "Valuing Managerial Flexibility". The Journal of Finance, Vol. XLV, No.2. June 1990.

[Trigeorgis and Mason, 87] L. Trigeorgis, J. Mason in "Valuing Managerial Flexibility". Midland Corporate Finance Journal, 5-1987. pp. 14-21.

[Trigeorgis, 96]   L. Trigeorgis. *Real Options: Managerial Flexibility and Strategy in Resource Allocation*. MIT Press, 1996.

[Wallace *et* al., 96] D. R. Wallace, M. Jakiela, W. C. Flowers. "Design Search Under Probabilistic Specifications Using Genetic Algorithms". Computer-Aided Design, Vol. 28, N0. 5, 1996. pp. 404 – 406.

[Wertz and Larson, 99] J. Wertz, W. Larson (Editors). *Space Mission Analysis and Design*. Microcosm Press and Kluwer Academic Publishers, 1999.

Chapter 3

[Allen, 97]   T. J. Allen. "Organizational Structure for Product Development". MIT Sloan, WP # 166-97.

[ARINC, 99]   "Resolution Cost Factors for Diminishing Manufacturing Sources and Material Shortages". Prepared for the Defense Micro-Electronics Activity (DMEA), McClellan AFB, ARINC. 1999.

[ARINC, 00]   "Program Managers Handbook: Common Practices to Mitigate the Risk of Obsolescence". Prepared for the Defense Micro-Electronics Activity (DMEA), McClellan AFB. ARINC, 2000.

[Blanchard, 98]   B. S. Blanchard. *System Engineering and Management*. Wiley-Interscience, 1998.

[Borke *et* al., 98]   J. Borky, R. Lachenmair, J. Messing, A. Frink. "Architectures for Next Generation Military Avionics". Proceedings of the IEEE Aerospace Conference. Vol. 1, 1998. pp. 265-281.

[Evans *et* al., 2000] M. Evans, N. Hastings, B. Peacock. *Statistical Distributions*. J. Wiley, 2000.

[Fine, 98]   C. H. Fine. *Clockspeed: Winning Industry Control in the Age of Temporary Advantage*. Perseus Books, 1998.

[Fitzhugh, 98]   G. Fitzhugh. "Rapid Retargeting: A Solution to Electronic Systems Obsolescence". Proceedings of the IEEE 1998 National Aerospace and Electronic Conference, NAECON 1998. pp. 169-176.

[Fricke et al. 00]   E. Fricke, A. Schulz, S. Wenzel, H. Negele. "Design for Changeability of Integrated Systems within a Hyper-Competitive Environment", in INCOSE Colorado 2000 Conference, Denver, March 2000.

[GAO-01-288]   General Accounting Office. "Best Practices: Better Matching of Needs and Resources Will Lead to Better Weapon System Outcomes." GAO-01-288, March 2001.

[Geroski, 91]   P. A. Geroski. *Market Dynamics and Entry*. Blackwell, 1991.

[Hatch, 00]   S. Hatch. "Diminishing Manufacturing Sources and Material shortages Management Practices". Proceedings of the Diminishing Manufacturing Sources and Material Shortages (DMSMS) Conference, August 2000.

[Henderson and Clark, 91] R. M. Henderson, K. B. Clark. "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failures of Established Firms". Administrative Science Quarterly, 35 (1990): 9-30.

[Hitt and Schmidt, 98] E. Hitt, J. Schmidt. "Technology Obsolescence Impact on Future Costs". Proceedings of the 17th AIAA/IEEE/SAE Digital

	Avionics Systems Conference (DASC), Vol. I, 1998. pp. A33-1, A33-7.
[Iansiti, 98]	M. Iansiti. *Technology Integration*. Harvard Business School Press, Boston, 1998.
[IEEE, 98a]	IEEE Guide for Developing System Requirements Specifications. 1998 Edition.
[IEEE, 98b]	IEEE Standard for Application and Management of the Systems Engineering Process. IEEE Std: 1220-1998.
[Lippiz, 99]	M. Lippitz. "Technology Upgrade Management and Department of Defense Acquisition Reform". PhD thesis. Department of Engineering Economic Systems and Operations Research, Stanford University. May 1999.
[Luke *et* al., 98]	J. Luke, J. Bittorie, W. Cannon. D. Haldeman. "A Commercial Off-The-Shelf Based Replacement Strategy for Aging Avionics Computers". Proceedings of the IEEE 1998 National Aerospace and Electronic Conference, NAECON 1998. pp. 177-181.
[McKluskey and Das, 98]	F. McKluskey, D. Das. "Component Technology Insertion into Critical Avionics Applications". American Helicopter Society Annual Forum. May 1998. pp. 1159-1165.
[NASA 95]	NASA Systems Engineering Handbook, SP-6105, June 1995.
[Porter, 80]	M. E. Porter. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York, Free Press, 1980.
[Terborgh, 49]	G. Terborgh. *Dynamic Equipment Policy*. McGraw-Hill, 1949.
[Ulrich and Eppinger, 95]	K. Ulrich, S. Eppinger. *Product Design and Development*. McGraw-Hill, 1995.
[Walton, 99]	M. A. Walton. "Identifying the Impact of Modeling and Simulation in the Generation of System Level Requirements". MSc thesis. Massachusetts Institute of Technology. Department of Aeronautics and Astronautics. June 1999
[Wertz and Larson, 99]	J. Wertz, W. Larson (Editors). *Space Mission Analysis and Design*. Microcosm Press and Kluwer Academic Publishers, 1999.

[Requirements Management Guidebook, 96]   SWELT: RM0.3-9/30/96.

Chapter 4

[Agrawal, 86]   Agrawal, B. N., *Design of Geosynchronous Spacecraft*. 1st ed., Prentice-Hall, Englewood-Cliffs, NJ, , 1986.

[Babcock, *undated*]   Babcock, P. S., "An Introduction to Reliability Modeling of Fault-Tolerant Systems." Draper Laboratory. CSDL-R-1899.

[Dudley and Verniole, 97] Dudley, G., Verniolle, J., "Secondary Lithium Batteries for Spacecraft." ESA Bulletin Nr. 90: May 1997.

[Garrison et al., 95]   Garrison, T., Ince, M., Pizzicaroli, J., Swan, P., "IRIDIUM Constellation Dynamics – The Systems Engineering Trades." IAF 95-U.2.04. IAF Congress, Oct. 1995.

[Isakowitz et al., 99]   Isakowitz, S., Hopkins Jr, J. P., Hopkins, J. B.,. *International Reference Guide to Space Launch Systems*. 3rd ed., Reston, VA. AIAA 1999.

[Jilla, 99]   Jilla, C.,. "A Multidisciplinary Design Methodology for Distributed Satellite Systems." PhD dissertation Proposal. Department of Aeronautics and Astronautics, MIT, Cambridge, MA. 1999.

[Lamassoure and Hastings, 00] Lamassoure, E.,. Hastings, D., "Generalized Metrics for Optimization of Space Systems Cost-Effectiveness." AIAA Paper 2000-5323, Sept. 2000.

[Wertz and Larson, 99]   Wertz, J., Larson, W., (editors). *Space Mission Analysis and Design*. 3rd ed., Microcosm Press, Torrence, CA: Kluwer Academic Publishing, Dordrecht/Boston/London, 1999.

[Wertz and Larson, 96]   Wertz, J., Larson, J., (editors). *Reducing Space Mission Cost*. 1st ed., Microcosm Press, Torrence, CA: Kluwer Academic Publishing, Dordrecht/Boston/London, 1996.

Chapter 5

[Allen, 97]   T. Allen. "Organizational Structure for Product Development". MIT Sloan, Working Paper # 166-97.

[Amram and Kulatilaka, 99] M. Amram, N. Kulatilaka. *Real Options: Managing Strategic Investment in an Uncertain World*. Harvard Business School Press, Boston, MA, 1999.

[Avinger, 75]   R. Avinger Jr. "The Economics of durability". PhD dissertation, Duke University, 1968.

[Boden and Larson, 95] D. Boden, W. Larson. *Cost-Effective Space Mission Operations.* McGraw-Hill, New York, 1995.

[Bourdenave, 76]   P. Bourdenave. "Faut-il Allonger la Vie des Biens Durables?" Thèse effectuée au Commisariat General du Plan Equipememt et de la Productivité. Ecole Polytechnique de Paris, Janvier 1976.

[Conn, 78]   W. D. Conn. "Factors Affecting Product Lifetime". Report WSF/RA-780219, School of Architecture and Urban Planning, UCLA. August 1978.

[Fine, 98]   C. Fine. *Clockspeed: Winning Industry Control in the Age of Temporary Advantage.* Perseus Books, 1998.

[Henderson and Clark, 90] R. Henderson, K. Clark. "Architectural Innovation: The Reconfiguration of Existing Products Technologies and Failures of Established Firms". Administrative Science Quarterly, 35 (1990): 9-30.

[Maier and Rechtin, 00] M. Maier, E. Rechtin. *The Art of Systems Architecting.* 2nd Edition. CRC Press, 2000.

[OECD, 82]   "Product Durability and Product Life Extension: Their Contribution to Solid Waste Management". Organization for Economic Co-Operation and Development (OECD), Paris, 1982.

[Schmalansee, 70]   R. Schmalansee. "Regulation and Durability of Goods". Bell Journal of Economics and Management Science, 7, 1970. pp. 54–64.

[Ullrich and Eppinger, 95] K. Ullrich, S. Eppinger. *Product Design and Development.* McGraw-Hill, 1995.

[Wertz and Larson, 99] J. Wertz, W. Larson, (editors). *Space Mission Analysis and Design.* 3rd ed., Microcosm Press, Torrence, CA: Kluwer Academic Publishing, Dordrecht/Boston/London, 1999.

[de Weck, 01]   O. de Weck. "Multivariable Isoperformance Methodology for Precision Opto-Mechanical Systems". PhD thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 2001.

Chapter 6

[AIAA-G-042-1991] DFOSS Handbook. *Proposed Guide to Design for On-Orbit Spacecraft Servicing.* AIAA, 1991.

[Alkalai et al., 00]   L. Alkalai, S. Chau, K. Tso, A. Tai, W. Sanders. "GSU Middleware Architecture Design". Proceedings of the 5th IEEE International Symposium on High Assurance Systems Engineering (HASE 2000). Albuquerque, Nw Mexico, November 2000.

[Amram and Kulatilaka, 99]   M. Amram, N. Kulatilaka. *Real Options: Managing Strategic Investments in an Uncertain World*. Harvard Business School Press, 1999.

[Adams et al., 87]   R. Adams, A. Gross, C. Jennrich. "Remote Repair Demonstration of the Solar Maximum Main Electronic Box". First European In-Orbit Operations Technology Symposium. Darmstadt, Germany. ESA SP-272, November, 1987.

[Bernstein, 96]   P. Bernstein. *Against the Gods: The Remarkable Story of Risk*. Wiley and Sons, 1996.

[Brown and Peuter, 94]   D. Brown, W. de Peuter. "Satellite Servicing in GEO by a Robotic Service Vehicle". ESA bulletin, May 1994.

[Cook and Lindell, 99]   W. Cook, S. Lindell. "Autonomous Rendezvous and Docking for Future Spacecraft Missions". AIAA-99-4598.

[Davinic et al., 97]   N. Davinic, S. Chappie, A. Arkus, J. Greenberg. *Spacecraft Modular Architecture Design Study: Cost Benefit Analysis of On-Orbit Satellite Servicing*. IAA-97-IAA.1.4.07.

[Faulkner, 96]   T. Faulkner. *Applying Options Thinking to R&D Valuation*. Research-Technology Management, May-June 1996, 39(3).

[Hall and Papadopoulos, 99]   E. Hall, M. Papadopoulos. *GPS Structural Modifications for On-Orbit Servicing*. Space Technology Conference and Exposition, Albuquerque, NM. August 28–30, 1999. AIAA 99-4430.

[Hastings et al., 01]   D. E. Hastings, E. Lamassoure, A. Weigel, J. Saleh. *Policy Enablers for the Development of a Space-Based Infrastructure*. Proceedings of the Fifth Conference on Technology, Policy, and Innovation. Delft, Netherlands, June 26-29 2001.

[Kerstein et al., 94]   L. Kerstein, G. Degtyarenko. *A Joint Development and Operation of a Small Unmanned Service Vehicle*. International Astronautical Federation. Jerusalem, Israel. October 9–14, 1994. IAF-94-T.3.447.

[Lamassoure, 01]   E. Lamassoure. *A Framework to Account for Flexibility in Modeling the Value of On-Orbit Servicing for Space Systems*. MSc thesis. Department of Aeronautics and Astronatics, MIT, 2001.

[Leisman et al., 99]   G. Leisman, A. Wallen, S. Kramer, W. Murdock. *Analysis and Preliminary Design of On-Orbit Servicing Architectures for the*

	GPS Constellation. Space Technology Conference and Exposition, Albuquerque, NM. August 28–30, 1999. AIAA 99-4425.
[Matunaga, *et* al., 96]	S. Matunaga, Y. Fukushima, Y. Ohkami. *Concept and Analysis of Tugrobot for In-Orbit Servicing.* International Astronautical Federation. Beijing, China. October 7–11, 1996. IAF-96-A.6.02.
[Merton, 97]	R. Merton in his 1997 Nobel Lecture. Reprinted in *American Economic Review,* June 1998, 88(3), pp. 323-349.
[Neely and de Neufville, 00]	J. Neely, R. de Neufville. *Hybrid Real Options Valuation of Risky Product Development Projects.* Submitted to the International Journal of Technology, Policy, and Management. 2000.
[Polites, 99]	M. Polites. *Technology of Automated Rendezvous and Capture in Space.* Journal of Spacecraft and Rockets, 36(2): 280-291, March-April 1999.
[Reynerson, 99]	C. M. Reynerson. *Spacecraft Modular Architecture for On-Orbit Servicing.* AIAA-99-4473.
[SAMS, 88]	*Space Assembly, Maintenance and Servicing Study* (SAMS). Final report, Vol. I-IV. TRW No. SAMSS-196-199. July 1988.
[Trigeorgis, 96]	L. Trigeorgis. *Real Options: Managerial Flexibility and Strategy in Resource Allocation.* MIT Press, 1996.
[Waltz, 98]	D. M. Waltz. *On-Orbit Servicing of Space Systems.* Krieger Publishing Company, 1993.

Chapter 7

[Bernstein, 96]	P. Bernstein. *Against the Gods: The Remarkable Story of Risk.* Wiley and Sons, 1996.
[Kube and Zhang, 94]	C. R. Kube, H. Zhang. "Collective Robotics: From Social Insects to Robot"s. Adaptive Behavior Vol. 2, No. 2, 189-218. 1994.
[Parnas, 79]	D. Parnas. "Designing software for Ease of Extension and Contraction". IEEE Transactions on Software Engineering, Volume SE-5, Number 2, March 1979. pp. 128-138.
[de Weck, 01]	O. de Weck. "Multivariable Isoperformance Methodology for Precision Opto-Mechanical Systems". PhD thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 2001.

Appendix A

[Asimow, 62]   M. Asimow. *Introduction to Design*. Prentice-Hall, Englewood Cliffs, N.J., 1962.

[Becker, 73]   J. M. Becker. *A Structural Design Process Philosophy and Methodology*. PhD Thesis, University of California, Berkley, 1973.

[Bernstein, 00]   J. Bernstein General Exam. MIT. Department of Aeronautics and Astronautics. Date not available!

[Blanchard, 98]   B. S. Blanchard. *System Engineering and Management*. Wiley-Interscience, 1998.

[Fricke *et* al. 00] E. Fricke, A. Schulz, S. Wenzel, H. Negele. *Design for Changeability of Integrated Systems within a Hyper-Competitive Environment*, in INCOSE Colorado 2000 Conference, Denver, March 2000

[Harris, 75]   A. J. Harris. *Civil Engineering Considered as an Art*. Proceeding of Inst. of Civil Engineers, pp. 15-23. February 1975.

[Highsmith, 99]   J. A. Highsmith III. *Adaptive Software Development: A Collaborative Approach To Managing Complex Systems*. Dorset House Publishing, 1999.

[Hazelrigg, 96]   G. A. Hazelrigg. *Systems Engineering: An Approach to Information-Based Design*. Prentice-Hall, 1996.

[Hazelrigg, 99]   G. A. Hazelrigg. *An Axiomatic Framework for Engineering Design*. Journal of Mechanical Design, Vol. 121, September 1999.

[Iansiti, 98]   M. Iansiti. *Technology Integration*. Harvard Business School Press, Boston, 1998.

[Parnas, 79]   D. Parnas. *Designing Software for Ease of Extension and Contraction*. IEEE Transactions on Software Engineering, Volume SE-5, Number 2. March 1979.

[Suh, 90]   N. P. Suh. *Principles of Design*. Oxford University Press, 1990.

[Suh, 99]   N. P. Suh. *Axiomatic Design: Advances and Applications*. Oxford University Press. (to be published). Manuscript format, 1999.

[Walton, 99]   M. A. Walton. *Identifying the Impact of Modeling and Simulation in the Generation of System Level Requirements*.

MSc thesis. Massachusetts Institute of Technology. Department of Aeronautics and Astronautics. June 1999

[Wertz and Larson, 99]   J. Wertz, W. Larson (Editors). *Space Mission Analysis and Design*. Microcosm Press and Kluwer Academic Publishers, 1999.

International Council on Systems Engineering (INCOSE). *Systems Engineering Handbook*, 1998.

IEEE Standard for Application and Management of the Systems Engineering Process. IEEE Std: 1220-1998.

NASA Systems Engineering Handbook, SP-6105, June 1995.

Requirements Management Guidebook, SWELT: RM0.3-9/30/96.

215

Appendix A

Nomenclature

A.1 The Trilogy: System, Design, and Systems Engineering

In order to have a clear definition of Systems Engineering for instance, it is first important to understand what is meant by the word "System". Alternatively, an understanding of "Design" helps further define and refine the concept of a system, and vice-versa. A holistic approach to those three terms, "System", "Design", and "System Engineering" can hence be mutually beneficial. In the following, various definitions of the above three terms provided in the literature are reviewed.
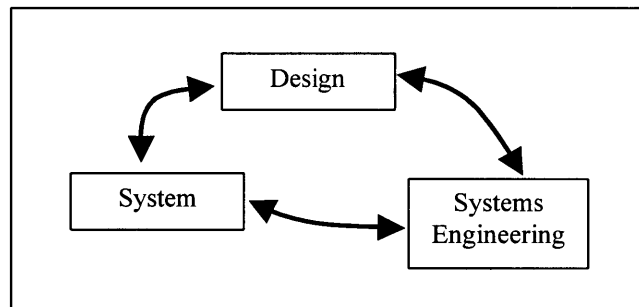
Fig. A-1. The Trilogy: System, Design, and Systems Engineering

A.1.1 What is a System?

From the Greek συστημα (systēma) meaning an organized whole, the word "system" is used in the common language to denote very different things. The first *behavioral* definition below provided by Ackoff presents the ideas of decomposition and interdependence of elements in a system and sets the stage for the more elaborate definitions to follow.

216

- Ackoff, 1974:

    A system is a set of two or more interrelated elements that must satisfy the following three conditions:

    (a) The properties or behavior of each element of the set has an effect on the properties or behavior of the set taken as a whole.

    (b) [...] No part has an independent effect on the whole and each is affected by at least one other part.

    (c) Every possible subgroup in the set has a non-independent effect on the whole. Therefore the whole cannot be decomposed into independent subsets. A system cannot be subdivided into independent subsystems.

- NASA Systems Engineering Handbook:

    A system is a set of interrelated components that interact with one another in an organized fashion toward a common purpose. The components of the system may be quite diverse, consisting of persons, organizations, procedures, software, equipments, and/or facilities.

    The following hierarchical sequence of terms for successively finer resolution was adopted by the NASA-wide System Engineering Working Group (SEWG): System, Segment, Element, Subsystem, Assembly, and Subassembly, part.

- IEEE Std 1233, 1998 Edition:

    An interdependent group of people, objects, and procedures constituted to achieve defined objectives or some operational role by performing specified functions. A complete system includes the associated equipment, facilities, material, computer programs, firmware, technical documentation, services, and personnel required for operations and support to the degree for self-sufficient use in its intended environment.

While the NASA and IEEE definitions of a system seem complete and overarching, I would simply add to them a *subjective* touch. By that I mean the following:

A difficult task in defining a system consists of specifying the system's boundaries and interfaces with its environment. It is likely that those boundaries are in the eyes of the person defining them,

and hence defining his/her system. For instance the Electric Power System on-board and aircraft with its Auxiliary Power Unit (APU), power regulators, electric actuators, power distribution, etc., can rightfully be called a *system* -and satisfies both the above definitions- from an electrical engineer's perspective. However from an aircraft designer perspective, this electric "system" is a subsystem like many others. Similarly from an airline executive's point of view, an aircraft is an element of a larger system that consists of the entire fleet, the people who operate it and maintain it, etc. One can continue this example from an FAA administrator point of view for instance, an airline is one component of a much larger system... This idea is discussed in the literature as systems that exist in the context of a broader *super-system* or a *system of systems*. The point of this paragraph is that a person's system's is another person's subsystem, and vice-versa, and that the definition of a system has to include the person formulating it.

A.1.2 Different Perspectives on Design

On could argue that the definition of "design" and that of a "system" are intrinsically related: design would be the activity that *creates* systems. One advantage for reviewing the literature on issues of design in addition to that of "systems" is that designers often propose definitions of **what** design is and **how** it should be conducted. Hence we get a first glimpse of the process that maps the customers' needs or identification of an opportunity into a system that satisfies those needs. This will also set the stage for the discussion of Systems Engineering where there is little agreement in the literature on its implementation approach (more on the "How" of SE than on the "What" is SE).

A significant amount of literature has been devoted to design. The works of Altshuller (1988, 1996), Asimow (1962), Becker (1973), Blanchard and Fabrycky (1990), Hazelrigg (1996), Sage (1977), Suh (1990), and Tribus (1969), and are some of recent texts that have addressed issues of design from both a philosophical and methodological perspective.

Design has been defined in a variety of ways depending on the specific context and/or field of interest. For instance, mechanical engineers often think of product design when they say or hear the word *design*, industrial engineers think of design in terms of new manufacturing processes and systems, entrepreneurs design organizations to achieve technical and/or financial goals, etc. All of the above involve design activities even though their content and the knowledge required to achieve the design goals are field-specific.

Suh (1999) defines design as an interplay or a mapping between **what** we want to achieve and **how** we want to achieve it:

> A rigorous design approach must begin with an explicit statement of "what we want to achieve" and end with a clear description of "how we will achieve it". Once we understand the customers needs, this understanding must be transformed into a set of specifications...that adequately describe the "what we want to achieve" to satisfy the customers' needs.

Intrinsic to Suh's Axiomatic Design framework is the concept of *domains* that distinguishes different kinds of design activities:

> The world of Design is made up of four domains: the customer domain, the functional domain, the physical domain, and the process domain (see below). The domain on the left relative to the domain on the right represents "what we want to achieve", whereas the domain on the right represents "how we propose to satisfy what is specified in the left domain".
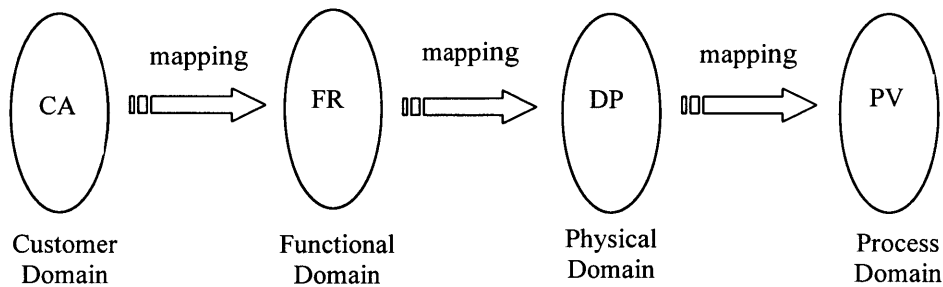

Fig.A-2. Four domains of the design world according to Axiomatic Design.

The customer domain is characterized by the needs or attributes (CA) that the customer is looking for in a product or process or service. In the functional domain, the customer needs are specified in terms of functional requirements (FR) and constraints. In order to satisfy the specified FR, the designer conceives of design parameters (DP) in the physical domain. Finally to produce the DPs, a process has to be developed characterized by process variables (PV) in the process domain.

Unlike Suh (1990, 1999), Hazelrigg (1996, 1999) restricts his discussion to engineering design. He builds upon the notion that design is a decision-making intensive process and develops a mathematics of design based on this idea and draws on theories from various fields such as

decision theory, utility theory, game theory, etc. The Accreditation Board for Engineering and Technology (ABET) also defines engineering design as a "decision-making process..." Hazelrigg's rationale is the following:

> Engineering involves the manipulation of nature to create systems for the benefit of at least some segment of mankind.

> This definition...implies that through the engineering [design] process, something physical is created...that has some value for at least someone. The process of creating something physical requires allocation of nature's resources; therefore engineering design is, essentially, the effective allocation of resources. The allocation of resources is by definition decision-making.

> [A] decision [is] an irrevocable allocation of resources. The selection of design parameters for an engineering system such as a computer or an automobile constitutes an allocation of resources. Design is a decision-making process, and the selection of design parameters represents decisions.

Asimow (1962) defines design as a "creative and purposeful activity directed toward the goal of fulfilling human needs". The words "creative" or "creativity" have been used in a variety of ways to describe the human activity that results in ingenious, unpredictable, or unforeseen results (e.g., new products, processes, and systems) that satisfy the needs of society or human aspirations. In this context, creative "solutions" are discovered or derived often without defining specifically what one sets out to create. "This creative spark may occur because the brain is, among other things, a huge information storage and processing device" [Asimow, 62]. Harris beautifully describes this stage of the design process:

> The designer...collects and assimilates as much facts as he can relevant to his design[4], using the full gamut of analytical techniques, if needed. He examines it, turns it over, changes it around, immerses himself in it, lets it sink into his sub-consciousness, drags it out again, walks around it, prods it. The hope is that, at some unsuspected moment, by who knows what mysterious process of imagination, intellect or inspiration, by influence

---

[4] This in Becker's model (1973) is called the "information base" which steadily grows and evolves as the design is refined.

220

of the genius, the daemon, the muse–the brilliant, the obvious definitive concept of the work will flash into the mind.

Several scientists have described making their discoveries in similar terms. However, a potential problem with this view for *design* is that it could reflect a lack of understanding of the process or the logic in the endeavor even though the result of the effort is intellectually, emotionally, or esthetically appealing.

A subject is always mysterious when it relies on an implicit thought process that cannot be stated explicitly and explained to others, and that can only be learned through experience, apprenticeship or trial-and-error. Recent attempts by authors such as Suh (1990, 1999) or Hazelrigg (1999) to develop an axiomatic[5] framework for design aim at making design a principle-based subject, hence reducing the random search and iterative trial-and-error process. Ultimately, their objective is to decrease the level of "mystery" in the design process as described above by Harris!

A.1.3 What is Systems Engineering?

I will restrict the discussion in this section to reviewing the various definitions of Systems Engineering that are available in the literature ("what"). The actual implementation approaches ("how") will be discussed in Part II.

- NASA Systems Engineering Handbook:

    Systems Engineering is an interdisciplinary approach encompassing the entire technical effort to evolve and verify an integrated and life cycle balanced set of system people, product, and process solutions that satisfy customer needs. Systems Engineering encompasses:

    (a) The technical efforts related to the development, manufacturing, verification, deployment, operations, support, disposal of, and user training for, system products and processes.

---

[5] "An axiom is a self-evident truth or fundamental truth for which there are no counter-examples. An axiom cannot be derived from other laws or principles of nature". In other words, "axioms are posited as accepted truths and a system of logic–framework–is built around them through the construction of theorems". This framework pervades broad classes of activities. These definitions are given by Suh (1990, 1999) and Hazelrigg (1999).

(b) The definition and management of the system configuration.

(c) The translation of the system definition into work breakdown structures.

(d) The development of information for management decision-making.

- The International Council on Systems Engineering (INCOSE):

  System engineering is the effective application of scientific and engineering efforts to transform an operational need into a defined system configuration through the top-down iterative process of requirement analysis, functional analysis and allocation, synthesis, design optimization, test and evaluation and validation[6].

- The DoD's MIL-STD-499 defines system engineering as the process that:

  (a) Transforms operational needs and requirements into an integrated system design solution through concurrent consideration of all life cycle needs, i.e., development, manufacturing, test and evaluation, verification, deployment, operations, support, training and disposal.

  (b) Ensures the compatibility, interoperability, and integration of all functional and physical interfaces and ensures that the system definition and design reflect the requirements for all system elements (hardware, software, facilities, people, data, etc.).

  (c) Characterizes and manages technical risks.

- Blanchard and Fabrycky define Systems Engineering as the process that involves the application of efforts to:

  (a) Transform an operational need into a description of system performance parameters and a preferred system configuration through the use of an iterative process of functional analysis, synthesis, optimization, definition, design, test, and evaluation.

  (b) Integrate related technical parameters and assure the compatibility of all physical, functional, and program interfaces in a manner that optimizes the total system definition and design.

  (c) Integrate performance, producibility, reliability, maintainability, and other specialties into the total engineering effort.

---

[6] International Council on Systems Engineering (INCOSE). *Systems Engineering Handbook*, 1998.

As can be seen from the definitions cited above, there is substantial overlap between them. They all describe for instance the systems engineering process as a **multidisciplinary, holistic,** and integrative approach that starts with the identification of operational or customers needs (How it proceeds is the subject of Part II). Another common theme in the definitions above is that Systems Engineering is both a management process as well as a technical process, and should include considerations of factors that a system might encounter throughout its **entire life cycle,** e.g., producibility, maintainability, supportability, and other –ilities.

The holistic approach to Systems Engineering can be better understood when contrasted with the traditional off-the-wall approach where once a task is completed, it is "thrown over the wall" to another team that will work on the next task. The holistic perspective in contrast suggests that all aspects of the design endeavor be considered simultaneously [Bernstein, 00]. The process is often described as an iterative one.

A.2 Other Definitions: Environment, Constraint, and Requirement

The followings definitions are those of the IEEE Std 1233, 1998 Edition.

A.2.1 Environment

The environment includes the circumstances, objects, and conditions that will influence the completed system. They include political, market, cultural, organizational, and physical influences as well as standards and policies that govern what the system must do or how it must do it.

A.2.2 Constraint

A Constraint is a statement that expresses measurable bounds for an element or function of the system. That is, a constraint is a factor that is imposed on the solution by force or compulsion and may limit or modify the design changes[7].

A.2.3 Requirements

A requirement is:

---

[7] Suh (1999) defines constraints as bounds on acceptable solutions. "There are two types of constraints: *input* constraints and *system* constraints. Input constraints are imposed as part of the design specifications. System constraints are constraints imposed by the system in which the design solution must function."

(a) A condition or capability needed by a user to solve a problem or achieve an objective.

(b) A condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed document.

(c) A documented representation of a condition or capability as in definition of (a) or (b).

Requirements can be taken from customer needs and can be derived from technical analysis.

- Capability (or functionality): Capabilities are the fundamental requirements of the system and represent the features or functions of the system needed or desired by the customer. A capability should usually be stated in such a way that it describes what the system must do. The capability should be stated in a way that is **solution independent**.

- Well-formed requirement: A statement of system functionality (a capability) that can be validated, and that must be met or possessed by a system to solve a customer problem or to achieve a customer objective, and is qualified by measurable conditions and bounded by constraints.

This definition provides a mean for distinguishing between **requirements** as **capabilities** and the **attributes** of those requirements. The following example is provided in the IEEE Guide for Developing System Requirements Specifications (1998) as a well-formed requirement:

Requirement: Move people from New York to California at a maximum speed of 5300 km/hr

Capability: Move people between California and New York
Attribute: Cruising speed of 2500 km/hr
Constraint: Maximum speed of 5300 km/hr