The Strategic Evolution of Systems: Principles and Framework with Applications to Space Communication Networks

by

Jennifer E. Manuse

S.B., Massachusetts Institute of Technology (2003)
S.M., Massachusetts Institute of Technology (2005)

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2009

Signature of Author . . . . . .
Department of Aeronautics and Astronautics
January 30, 2009

Certified by . . . . . .
Dr. Olivier L. de Weck
Associate Professor of Aeronautics and Astronautics & Engineering Systems
Thesis Supervisor

Certified by . . . . . .
Dr. Dorothy C. Poppe
Principle Member of the Technical Staff, The Charles Stark Draper Laboratory, Inc.
Committee Member

Certified by . . . . . .
Dr. Annalisa Weigel
Assistant Professor of Aeronautics and Astronautics & Engineering Systems
Committee Member

Certified by . . . . . .
William D. Ivancic
NASA Glenn Research Center
Committee Member

Accepted by . . . . . .
David L. Darmofal
Associate Department Head
Chair, Committee on Graduate Students

The Strategic Evolution of Systems: Principles and Framework with Applications to Space Communication Networks

by

Jennifer E. Manuse

Submitted to the Department of Aeronautics and Astronautics on January 30, 2009, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Aeronautics and Astronautics

Abstract

Complex systems in operation are constrained by legacy; in other words, the existing properties, structure, components, protocols, software, people and etc. that are inherited over time. This inheritance heavily influences the type and timing of feasible and available changes to the system, the effectiveness and desirability of their impact while accounting for uncertainty, and the future constraints imposed as a result. This thesis introduces the Strategic Evolution of Systems, a novel framework for evolving complex systems that directly addresses legacy challenges during system operation within the context of space communication networks. The framework — perspective, position, plan and pattern — is based on Mintzberg's "emergent" interpretation of strategy.

This thesis also presents several unique ideas including the concept of option lock-out, or the tendency to lose access to potentially desirable regions of the architectural space when exercising a transition; an energy analogy to model static architecture value; an entropy-based formulation to evaluate the desirability, or dynamic multidimensional value, of an architecture by considering the structural and temporal space of possible transitions; and the application of the entropy-based formulation to define the overall desirability of an architecture as its position, or current situation (favorable or unfavorable) relative to accessible alternatives, in order to identify the most advantageous immediate transition.

A key contribution of this thesis is a method to value legacy in a physical non-market traded system, including a demonstration of its application to a system in which benefits and costs are non-monetary in nature. Other important contributions include a change exposure tool, referred to as a Strategic Advantage Map, to visualize the near- and long-term impact of immediate transitions relative to legacy. Here, an architecture's position relative to the legacy system can be thought of as the region of entropy space it occupies (evaluated over time and uncertainty). The more dominating this region of position entropy is, the more desirable the architecture. For monetary-based systems, a second change exposure tool includes an "Iceberg Exposure,"which maps the exposure of net present value for each accessible transition option relative to a neutral no-gain-no-loss line, resulting in a graph resembling an iceberg. The visualization tools allow decision makers to quickly evaluate the impact (risk/opportunity) of change, based on their concept of desirability.

Case studies include a historical look at the NASA Deep Space Network for insight into legacy and complex system evolution, a demonstration of the Strategic Evolution of Systems framework for a global commercial satellite communication system, and an illustration of the method extended to non-monetary systems for the deployment of communication assets to support manned exploration

of Earth's moon. The satellite system case study introduces an extended market model that evaluates the attainable business segments in a global satellite communications system by integrating estimates of the global distribution of market demand, observed traffic statistics, and calculations of the resulting steady-state network performance.

This thesis will show how to use the framework and principles for evaluating a system's current position as well as how to update the evaluation as time progresses. The satellite communication case study will provide one example where the methodology enables identification of the optimal transition path over the system's operational life. It will become evident that the choice of horizon time and the use of debiasing factors can have significant influence on the results. Future study on properly identifying and constructing these variables is strongly recommended. Finally, the ideas and tools presented in this thesis may be used to compare preferred systems to suggested alternatives in order to justify expenditures or to initiate research and development programs.

ACKNOWLEDGMENT

30 January 2009

This thesis was prepared at The Charles Stark Draper Laboratory, Inc., under the Draper IRAD #20349-001, NASA SAIR SE #21045-002, Draper Fellows #22026-001 and the Draper Fellow Slush Fund #22951-001.

Publication of this thesis does not constitute approval by Draper or the sponsoring agency of the findings or conclusions contained therin. It is published for the exchange and stimulation of ideas.

PERSONAL ACKNOWLEDGMENTS

For my beautiful, beloved Aidan Rose.

I could not have completed my thesis work without the love and unwavering support of my dear husband, Andrew. He believed in me even when I did not, and he encouraged me to push the boundaries of thought and sleep so that I could achieve my goal of finishing before Aidan's second birthday. Thank you for being my husband, my best friend, the father of our daugher and the man who elevates me.

I owe a tremendous amount of gratitude to Dr. Dorothy Poppe, my Draper supervisor, who was bound and determined that I would succeed and overcome the numerous unexpected challenges that have arisen during the last few years. Even during the year and a half she spent away to pursue a prestigious fellowship at the NRO in Virginia, she worked hard to support me. She is my mentor, my advocate and my friend, and I look forward to working with her in the future.

I want to thank the other members of my Committee, Dr. Olivier L. de Weck, Dr. Annalisa Weigel and William Ivancic for their many insights and ideas that have helped shape the work presented here. It has been a pleasure working with you and I hope we have the opportunity to work together again.

To my parents, Linda and Steve Underwood: it has been a long road to get here, but I did it! I hope the success of my defense on your wedding anniversary was a special gift. Thank you for always being there for me and encouraging my talents. I wouldn't be here if it weren't for everything you have given me over the years.

To James and Nancy Manuse: thank you for welcoming me into your family and supporting me in every possible way. I am very fortunate to have in-laws as wonderful as you.

Contents

List of Figures

List of Tables

Nomenclature

α	A debiasing factor for the effect of the unit fraction in the configuration term.
β	Decision-maker defined weightings.
ΔG	Change in Gibbs free energy.
ΔH	Change in state enthalpy.
ΔS	Change in entropy.
Δt	A discrete interval of time.
γ	A debiasing factor to reduce the influence of the number of artificially-scoped option paths.
λ^{SD}	Packet arrival rates between a source S and a destination D.
λ^{S}	Packet arrival rate to each source node.
Ω	The number of microstates corresponding to the observed thermodynamic macrostate; 'the ratio between the number of possible (spatial) arrangements of system components that can give the current configuration and the total number of possible arrangements yielding all possible configurations of the system.'
ω	A dummy variable to represent a probability.
σ	Volatility [%].
B_B	The benefits, or value-delivery, anticipated for the benchmark architecture.
$B_i(t)$	Benefit accrual for architecture i until time t.

$B_T(i \rightarrow j)$ The benefit of operating an architecture i over the time period GIVEN the architecture is being transitioned from architecture i to j.

C_B The costs, or resources expended, to achieve the benchmark architecture.

$C_i(t)$ Costs incurred, or resources expended, by architecture i until time t.

$C_r(t)$ Costs incurred to retire an architecture at time t.

$C_T(i \rightarrow j)$ The cost of operating an architecture i over the time period AND the cost of transitioning between architectures i and j.

$C_{BN}(j,t)$ Costs incurred from building architecture j during the period (t-1,t].

$C_{RO}(i,t)$ Costs incurred from ripping out architecture i at time t.

E_B The energy of the benchmark architecture.

h_i A dummy variable to represent system desirability.

H_T The transition entropy associated with transitioning between the current configuration and a new configuration.

k_B Boltzmann's constant.

p_i The probability that particle i will be in a given microstate; also, the probability that a given transition option is exercised.

$p_{i \rightarrow j}$ The probability that a decision maker will chose to transition the current architecture i to a new architecture j.

V_j The value of architecture j, defined here as the inverse of its energy.

AddC Add lunar satellite constellation plane decision.

AddLAN Add excursion Local Area Network decision.

ARQ Automatic repeat request protocol.

B Benefit, or value-delivery, of an architecture.

C Cost, or resources expended to achieve a given level of benefits.

C(k) The cardinality (number of elements) of the set of options k.

CAlt Constellation altitude at semi-major axis.

CEcc Constellation orbital eccentricity.

CIncl Constellation orbital inclination.

Cper Constellation periapsis.

CRaan Constellation right ascension of the ascending node.

Deploy Relay deployment strategy.

Dr Receiver diameter.

Dt Satellite transmitter diameter.

E System energy.

Emin Minimum elevation angle.

H The position entropy of a system.

h Orbital altitude.

H(X) The information entropy H of a discrete random variable X.

k The set of immediately accessible options.

LANType Type of LAN product platform.

MAP Multiple access protocol.

MS Modulation scheme.

ND Network routing architecture decision (centralized vs. distributed).

PS Average data packet size.

Pt Satellite transmitter power.

RP Routing protocol.

Ruser Average user data rate.

S	Thermodynamic entropy.
T	Temperature.
t	Time.
TC	Terrestrial capacity.

Chapter 1

Introduction

Complex systems in operation are constrained by legacy; in other words, the existing properties, structure, components, protocols, software, people and etc. that are inherited over time. This inheritance heavily influences the type and timing of feasible and available changes to the system, the effectiveness and desirability of their impact while accounting for uncertainty, and the future constraints imposed as a result. This thesis presents the Strategic Evolution of Systems, a framework and principles for evolving complex systems that directly addresses legacy challenges during system operation within the context of space communication networks.

This introductory chapter serves to provide motivation for the thesis by way of simple examples, to further refine the thesis objectives and approach, and to lay the groundwork for the development of the Strategic Evolution of Systems theory by providing the necessary background, literature review, and concept development. This chapter will conclude with a detailed list of thesis contributions and an overview of the rest of the thesis.

1.1 Motivation

To motivate this thesis, consider a simple example in which a design team is tasked with choosing an avionics processor architecture. It shall initially be assumed that the design decisions are made independent of time. When this assumption is relaxed, the intricacies involved with making decisions over a time horizon are discussed. It will be shown that while the choice of optimal build architecture is clear when looking at an instant in time, the optimal decision is considerably less knowable

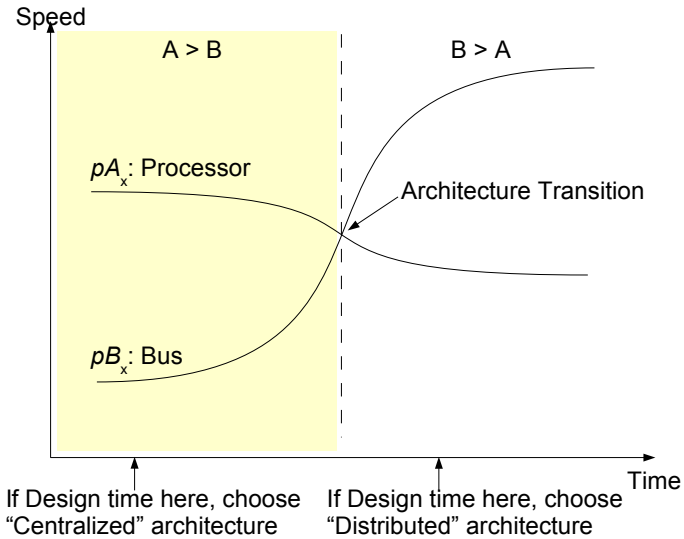


Figure 1.1: A simple example of the time-varying nature of optimal build decisions and the architecture transition point that separates them for an avionics processor with two parameters: bus and processor speed.

when the effect of time is introduced. This difficulty arises due to the existence of legacy. The potential issues arising with legacy migration are addressed in the discussion of transition parameters and their evolution over time.

One of the avionic processor architecture team's major design decisions would be whether the avionics system should have a distributed or centralized architecture. If cost is held constant, there are two key design tensions driving the decision: processor speed and bus speed.[1] The application for the avionics architecture would have requirements concerning the minimum values for these parameters.

If the decision is made independent of time, then the appropriate design decision would be dependent on the relative values of the parameters: "centralized" if the processor speed is greater than the bus speed, and "distributed" if the bus speed is greater than the processor speed. But what happens when time is incorporated?

As can be seen in Figure 1.1 the relative values of the two speed parameters are always changing, and the regions of ideal decisions are clearly marked. There will be times when they flip positions. In the design phase, it is important to account for this movement. If the design time falls safely within the first region, well before the switch is anticipated, then the ideal system would likely be

[1] The bus transfers data between the processor that interprets instructions and the other avionics computer components.

"centralized"; similarly, the ideal system would likely be "distributed" if the design time falls safely within the second region.

There is a caveat here since there are lag times between the decision and manufacturing and implementation phases, and the system would have an expected life time for operational purposes. If the switch is expected to occur before the manufacturing phase, the ideal system would now be "distributed," but it might be too late to change the architecture without incurring significant cost and time penalties. This effect indicates that if the design time falls on or near the transition point between the two architectures, then the design team would need to spend time examining the forecasts to decide which architecture is most likely to be ideal. The lag between a system decision and manufacturing of that system could easily result in a switch of the ideal architecture.

Up until this point, there has been no mention of minimum requirements for the speed of the processor and bus architecture. Depending on the purpose of the hypothetical avionics processor, these requirements may or may not be expected to change over time. If the requirements are likely to be static, then the requirements will be the main driver of the ideal architecture chosen in the design phase. It is possible that any combination of bus and processor speed cannot meet the full requirements under the current conditions, but that does not mean this will always be the case.

If the full requirements are not relevant until a later time in the operation, then it may be possible to construct the system such that there is a real option for substituting in the new ideal architecture when the time comes for full operational capability. With an integrated component like an avionics processor system, this could mean significant integration and test costs, particularly since the component is so strongly coupled to many other subsystems.

An even more interesting case is when the requirements are expected to change over time but in an unknown fashion. The design team can incorporate real options into the design at an immediate penalty, but the real challenge comes during the operation of the system. Now, there is a legacy system (and many times, there have not been real options intentionally built into the system), there might or might not be the option to take the system offline for upgrades, components are no longer operating at their optimal, and maintaining the system without modification runs a real chance of doing more harm than good to the system and system performance.

A real world example of the existence of transition parameters and their shifting effect on the optimal build architecture is described in *Nuclear Power Reactors: A Study in Technological Lock-in.*[13]

> *Important in the competition between direct and alternating electric current were changes in the environment in which technologies operated, altering which aspects of the technology were desirable and which ought to be considered "best." Initially, when electricity transmission took place over short distances, direct current seemed to have an advantage, as it was technically better equipped to meet peak loading problems. As transmission distances grew, however, alternating current gained the upper hand, to a great extent due to its ability to transmit at high voltages and then use step-down transformers to lower the voltage for consumers, thus circumventing the problems of voltage loss. Now, however, direct current is making somewhat of a comeback, with very long-distance, very high-voltage transmission.*

As can be seen from the processor architecture example and the nuclear power reactor example, understanding and valuing legacy is important, both in the design and operation of systems.

1.2 Thesis Objectives and Approach

This section outlines the thesis objectives and the approach used to meet them.

As discussed in the introduction of this chapter, the primary objective of this thesis is to develop a framework that directly addresses legacy challenges during system operation within the context of space communication networks. The first challenge is that legacy influences what changes are feasible and available as well as when a change may be made and how long it may take to implement. The second challenge is that legacy influences the effectiveness and desirability of any change that is made while accounting for uncertainty. Finally, legacy influences the future constraints placed on the system as the result of making further changes.

The high-level question to be answered is: accounting for uncertainty, if a decision maker desires to make a change to an operating system, what is the most advantageous option to exercise and when should it be exercised? In other words, how can decision makers strategically evolve systems over time in the face uncertainties both within a system and outside of it?

As will be seen in Section 1.3.3, advantage is defined as "to be in a superior or advantageous position," and the concept will later be modified to evaluate the desirability of making a change.

The primary thesis objective can be broken down into a series of questions that will be answered throughout the thesis.

First challenge:

- What options are feasible and available to exercise?
- How can legacy contraints on when a change may be made and how long it may take to implement be modeled and accounted for?

Second challenge:

- What immediate migration of the system will achieve the overall or long-term advantage while still meeting the needs of today?
- When, if at all, should this option be exercised?

Third challenge:

- Can the future constraints on a system be identified, and, if so, can their effect be estimated?
- Can decision makers avoid locking the system into a bad architecture?
- At what point should the system be ripped out and redone rather than relying on legacy migration?

This thesis also has several secondary, yet related objectives.

- To ascertain a set of principles based on historical examples that can guide decision makers.
- To find a method to value a physical non-market traded legacy system.
- To find a method to evaluate desirability when the benefits gained and resources expended are non-monetary in nature.
- To find a way to visualize the near- and long-term advantage.

In order to meet the desired thesis objectives, the following approach was undertaken.

- **Observe:** the transition from IPv4 to IPv6, the NASA Deep Space Network history and evolution.
- **Identify:** principles, patterns and insights.
- **Create:** theory, metrics, tools, framework.
- **Apply:** the global commercial satellite communication system case study and the deployable communication networks for manned lunar exploration case study.

As will be seen at the end of the chapter, the organization of the thesis will follow this approach.

1.3 Background and Literature Survey

This section provides the background and literature review required to lay the groundwork for the development of the Strategic Evolution of Systems.

One of the important distinctions of this research is that it assumes a system in operation with legacy. The big question for a decision-maker becomes: "How do I make the most with what I have?" The decision maker must understand the nature of the legacy, the decision maker must frame and implement a strategy in order to achieve the desired result, and the decision maker must understand the context in which decisions are made.

1.3.1 The Two Faces of Legacy

Legacy simply refers to the existing system and the fact that its properties, components, protocols, software, people and etc., are inherited over time. Any changes to the system are constrained by this inheritance. Components can be removed and protocols changed, but there are immediate consequences to these changes because of the structural, functional and procedural relationships between these components and the rest of the legacy system. These consequences are not always predictable and can be quite devastating to the system. For example, the switch to a new insulating foam on the space shuttle boosters indirectly led to the Columbia shuttle disaster.

The impact of legacy on information systems and communication networks is becoming more apparent. In the past, these systems focused on the future application of Moore's law and ignored old systems as they became obsolete. However, this strategy is quickly becoming ineffective as these systems become more and more interconnected with others outside the organization, which may be under the purview of another decision maker[14]. Interoperability and backwards compatibility are therefore becoming more important.

The effect of the "inertia" of legacy is quite apparent when considering the current Internet Protocol transition from IPv4 to IPv6. IPv4 has been used for more than 30 years and IPv6 is only now maturing. Although the U.S. Department of Defense has mandated the transition with core capability by 2008 for their systems, and the United States (U.S.) government has since followed suit with its

agencies, businesses are reluctant to make the switch because they do not seem to believe there is a threat of depleting address spaces. Businesses perceive that there are more cost-effective ways to minimize this effect via the use of Network Address Translation (NAT) and/or proxies.

CISCO and Juniper have been adding IPv6 functionality to their routers since 2001 and Microsoft is now supporting IPV6 and Next Generation Transmission Control Protocol/Internet Protocol (TCP/IP) stack in their emerging Windows products. With the ability to run IPv4 over IPv6, in combination with NAT and proxy address space fixes, it may be some time before businesses decide to make the transition.

This thesis proposes that there are not one but two faces of legacy. The first is well known and well researched in a variety of disciplines: Architectural lock-in (also known as technological lock-in) reduces a systems' ability to adapt or to be changed by external forces. This thesis proposes a second perspective to legacy: Option lock-out, which is related to the path-dependent nature of many systems, but is not a direct measure of path dependency[15] nor merely of the reduction in the degrees of freedom.[16] As will be shown, the first is inward looking, the second is forward looking.

Architectural Lock-in

Architectural lock-in can be thought of as the tendency to maintain a system "as is," to keep the status quo, and the resistance of decision makers to change the system. This artifact of legacy systems reflects risk-averse behavior. The resistance to change legacy systems conjures memories of physics class and discussions of inertia. So, in a sense, architectural lock-in behaves as a kind of legacy inertia, making the system hard to move around and change according to its circumstances.

Silver[15] provides a very nice overview of architectural lock-in as well as its relationship to path dependency. A relevant example cited is the ongoing deliberation over whether NASA should continue with Shuttle-derived vehicle architectures or if a switch should be made to the Extended Expendable Launch Vehicle (EELV)-derived Heavy Lift Launch Vehicle (HLLV) systems. Both architectures appear to have the necessary capability, but the incentives favor the Shuttle-derived option. Although space systems suffer from significant complexity and high capital cost, much of the resulting risk and cost for the Shuttle have been retired. Developing an architecture based on EELV components carries a certain amount of uncertainty that is not found in Shuttle systems. Political motivations come into play as well, since the preference is to maintain jobs within existing programs.

Silver argues that the fundamental, multidimensional causes of lock-in (and therefore path dependency) can be captured within the concept of a switching-cost for moving between two architectures. He discusses the effects of sunk costs, sunk cost hysteresis, network effects, learning and culture.

The idea of a sunk cost has already been put forward in the previous discussion. A sunk cost refers to costs, risks and learning curves that have been retired in the course of the development, implementation and operation of the legacy system, regardless of whether the technology that has been implemented is suboptimal or inefficient. Thus, there is an inherent incentive to put off changing the system. This tendency increases, Silver argues, when the operating environment is uncertain and the investment is irreversible.

Waiting for uncertainties to resolve seems the better choice since the investment in changes cannot be reversed. For a new system to be the preferred choice, the total projected expenses must exceed a "hurdle rate" that accounts for the uncertainty. Economists have defined "sunk cost hysteresis" as the effect of the hurdle rate on investment.

Network effects — the increase in value of certain technologies as more people use them — amplify switching costs. "The benefit of a technology stems from its relationship to other processes and needs, the totality of which defines a value network." Silver points to the metric system, the qwerty keyboard and Video Home System (VHS) tapes as examples of this phenomenon. As the technology became more established within the culture, the value of the technologies increased with the increasing number of end users as well as developers of other technologies. As more people used the technology, new products were developed that incorporated the technology or were built to interface with the technology. If the core technology (metric system, qwerty keyboard, and etc.) changes, an ever-increasing number of other technologies coupled to its use must also be changed. Thus, there is significant legacy inertia associated with systems exhibiting network effects, which can also be thought of as de facto standards that define interfaces between components of its value network.

Learning and culture is the final component of architecture lock-in addressed by Silver. Any changes to a system requires "paying down" the learning curve of the original system again. Similarly, engineers tend to reuse solutions to past problems, which can influence the "nature" of new problems. A change to the system must also overcome concentrations of expertise in areas that may not be as relevant to the new system or for the transition. The change must fight political and organizational inertia in order to complete any necessary restructuring of the organization, procedures and/or lines

of communication among team members and groups.

Option Lock-out

Silver[15] describes path dependence as "the fact that future options are constrained by past decisions." Path dependence is often caused by architectural lock-in. It should be noted that path dependence is value-neutral. There is no mention in the description of path dependence of whether the future options are desirable or what potential options would be lost in the process of making further decisions. Nor are either of these important questions addressed within the definition of architecture lock-in.

Architectural lock-in views legacy as a hindrance toward evolving a system strategically. Legacy is a limitation, but there is also enormous value inherent in the resources already invested. This value needs to be acknowledged, identified and leveraged. Legacy Value Engineering[17] reflects this philosophy, but it is not the only research into valuing legacy and evaluating how best to leverage its value. The Legacy Systems Engineering project in Arizona State University's Department of Mechanical and Aerospace Engineering[18] has been working to maximize the value of long-term legacy military systems. Neither of these efforts appear to address the desirability of future options or what future options have been lost because of past decisions. Nor do they address what might be lost during the process of making near-term future decisions.

Thus, access to the space of options appears to be missing from the studied notion of legacy and its evolution.

Define **option lock-out** as the tendency to lose access to potentially desirable regions of the architectural space. In other words, once a system is changed, there are some architectural possibilities that were possible and desirable before the change but are either no longer possible or desirable because of the change. As an example, consider the various uses of eggs. Once an egg is boiled, it is no longer possible to hatch a chicken or make cake batter from that egg.

To illustrate the difference between architectural lock-in and option lock-out, consider the nuclear reactor example Silver uses to motivate architectural lock-in.[15]

> *Light water is considered inferior to other technologies, yet it dominates the market for power reactors. This is largely due to the early adoption and heavy development by the U.S. Navy of light water for submarine propulsion. When a market for civilian power*

emerged, light water had a large head start, and by the time other technologies were ready to enter the market, light water was entrenched.

It turns out that while it is still technically possible to transition from light water to other technologies by building the desired type of reactor rather than a light water version, human, political, and influence factors pressure against the transition, leaving the choice of technology for reactors architecturally locked-in during the operation of the reactor. The options are still technically feasible, so option lock-out has not taken place. However, if one looks at a single reactor that is currently in operation, there is considerable doubt whether it would be possible to transition from light water to another technology without ripping out the structure and building from scratch. In this case, not only is there architectural lock-in on the basis of the previous argument, but now there is option lock-out as well. The reactor is effectively locked-out of the option to switch technologies during operation because it is too expensive and time consuming to rip out and build new.

1.3.2 Strategy

The concept of strategy is at least as old as ancient Greece. The Greek word *strategia* means "generalship", and its origin is clearly military in nature. Since then, the philosophy of strategy has evolved along with its application. Recently, military strategy was ported to the business world, where troops were replaced with system-specific resources.

At the highest level, strategy links policy with tactics. A decision maker identifies the desired end result and uses strategy and tactics to identify the best means to the end (there are numerous ways to define strategy).

The following definitions are from a collection of definitions for strategy compiled by Fred Nickols.[19]

Liddell Hart describes military strategy as "The art of distributing and applying military means to fulfill the ends of policy."

Henry Mintzberg, in commenting on the rise and fall of strategic planning, summarizes the four most common uses for strategy. He integrates these uses and proposes a "realized" strategy:

- Strategy is a plan, a "how," a means of getting from here to there.

- Strategy is a pattern in actions over time. For example, a company that regularly markets very expensive products is using a "high-end" strategy.
- Strategy is a position; that is, it reflects decisions to offer particular products or services in particular markets.
- Strategy is perspective; that is, vision and direction.

According to Nickols[19]:

> *Mintzberg argues that strategy emerges over time as intentions collide with and accommodate a changing reality. Thus, one might start with a perspective and conclude that it calls for a certain position, which is to be achieved by way of a carefully crafted plan, with the eventual outcome and strategy reflected in a pattern evident in decisions and actions over time.*

Essentially, strategy is concerned only with how to achieve aims, not what those aims are or ought to be.

Strategic Engineering has recently been defined as "the process of architecting and designing complex systems and products in a way that deliberately accounts for future uncertainty and context in order to minimize the effects of lock-in while maximizing life-cycle value."[20]

This thesis proposes an extension to the concept of strategic engineering that acknowledges the art of distributing and applying resources to fulfill the system objectives during the operation cycle.

Adapting and reordering the four main components of the Mintzberg definition of strategy to include the engineering of complex systems gives:

- **Perspective:** Determination of the relevant information, how this information is viewed and specification of the methods used to evaluate the information.
- **Position:** Current situation (favorable or unfavorable) relative to alternatives.
- **Plan:** Guidelines for taking the position within the context of perspective and making an actionable decision.
- **Pattern:** Observations of the actual decisions by decision makers over time.

1.3.3 Advantage

It is reasonable to assume that every engineering system, organization, business, and etc., aims to be in a superior or advantageous position. This goal happens to be the definition of advantage.[21]

A position is arguably superior if it has more value over some time horizon. The difficulty with defining value is that it is perceived in the eye of a beholder. An entire field of ongoing research has resulted in numerous methods and proposed quantifications. Several of these definitions are described below.

Value in exchange is the monetary value of a system, which could be estimated as the market price (i.e., monetary value if available),[22] or the "as is" replacement cost. Value in use is a utility measure of value by way of an estimation of the nature (i.e., essential characteristics), quality (i.e., grade of excellence), extent (i.e., degree that purpose is achieved) and/or significance (i.e., importance relative to system as a whole) of the system in question.[22] This is a static definition (i.e, choose a moment frozen in time and estimate the value).

Hitchins[23] argues that in order to establish a valid measure of value, there are three things that must be done. First, "establish it in a systems context, interacting with other, complementary systems in a containing system." Second, "establish the purpose of the containing system." Third, "establish the degree to which the something contributes, along with its siblings, to the containing system's purpose." These steps mirror the extent and significance aspects of value in use. Hitchins describes an interesting application of the method:

> *For a hospital, the containing system is the area or region served by the hospital, with its fabric of management. The containing system's objective might be to maintain, enhance, and restore the quality of life of the residents. Sibling systems are those are interact with each other to contribute to that objective, and they will include environmental systems, waste disposal systems, education systems, transport and communication systems, sport and leisure systems, policing, and so on.*
>
> *The value of the hospital can now be seen in context: both in terms of its tangible impact on the quality of life compared with other sibling systems, and in terms of its psychological impact on the quality of life by assuring the society, including those with no immediate need of its services, that it is there, ready, and able to perform should unforeseen difficulties arise.*

> *Note, too, that the various siblings interact. The risk from poor waste disposal is disease. The lack of sport and leisure facilities may lead to heart disease, drug taking, drunkenness, crime, unrest, and disturbance, with injuries to abusers, criminals, rioters and police. This increases the hospital's customers, not to mention the need for more police, and there is a consequent reduction in the amounts of money available for other activities.*
>
> *To assess the value of a hospital sensibly would require a dynamic model of the society with its various interacting systems, showing how they individually and together contribute to maintaining, enhancing, and restoring the quality of life, measured perhaps by some index. Changing the subsystems in the model would change their interactions and the overall quality of life index. In this way it would be possible to establish not only the degree of contribution for each and every subsystem, but also how they might be changed so as to increase the index overall.*

Much of the value related to evolving complex systems is derived from the value generated by the transitions themselves. Often this value is correlated to the reasons for the change.

Fricke and Schulz[24] argue that the "steady insertion of new technologies is necessary" for system competitiveness. Changes are often made to save money in the long run, to increase the quality of the product, to return to a planned schedule, to correct incomplete or flawed requirements or to meet changing requirements. Changes made for these reasons create value for the system undergoing transition.

Thus, the strategic advantage in operation can be defined in terms of being in a superior or advantageous position over some time horizon. This definition implies that advantage behaves as a kind of dynamic value.

Advantageous Position for Systems:

- **Near term:** Accessible architecture has greatest forecasted Value-Opportunity-Risk (VOR) over the valid forecast horizon (e.g., one transition period).
- **Long Term:** Accessible architecture has greatest access to future architectures (in terms of number of options and desirability of those options).

Suppose an accessible architecture has both the immediate advantage and the long-term advantage over other architectures. The position of this accessible architecture whould be superior to other

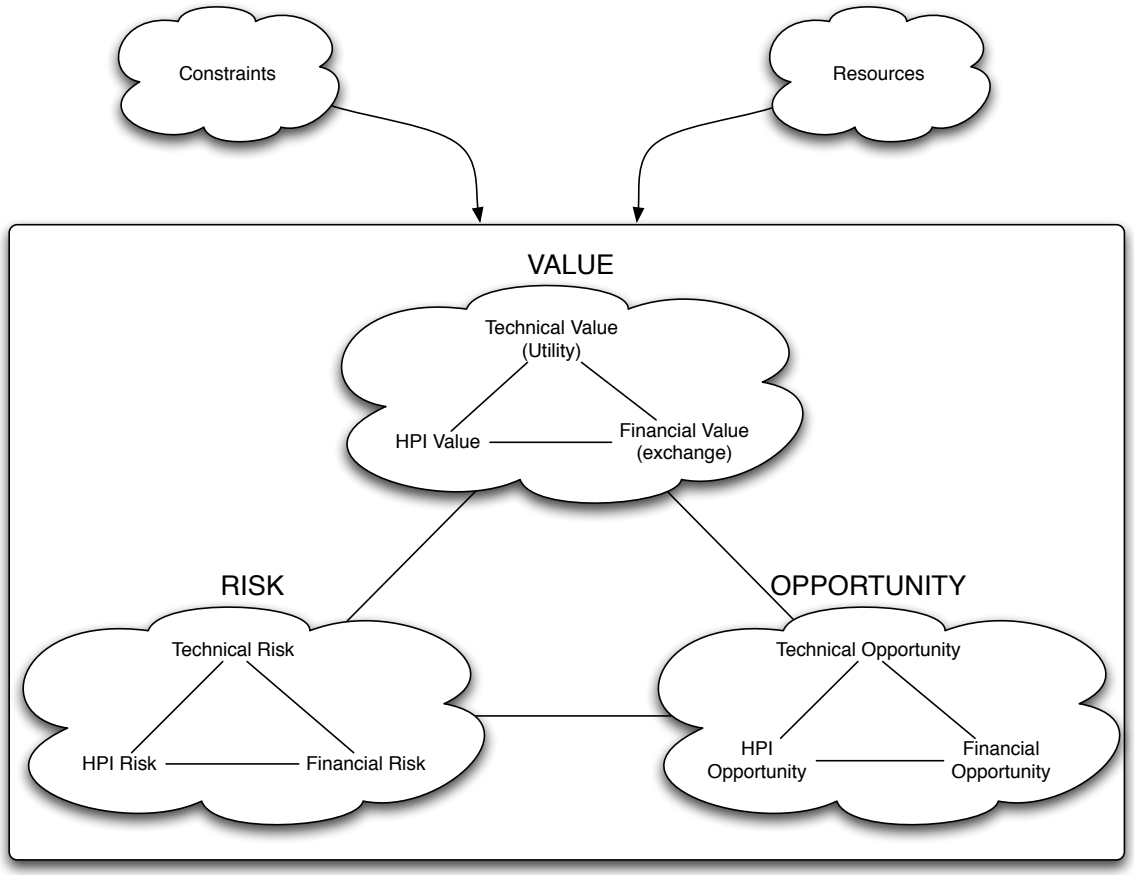


Figure 1.2: The components of advantage: value, opportunity and risk (VOR). The diagram depicts the coupled interactions between technical VOR, financial VOR and the VOR associated with the nature of human resources, politics and influence (HPI). These components must work together to achieve the "optimal" value of the system.

architectures. However, if only one or the other advantage is true, then the superiority of the position of this accessible architecture is less clear.

The relationship among value, opportunity and risk — the components of advantage — are shown in Figure 1.2. The diagram depicts the coupled interactions between technical value, financial value and the value attached to the nature of human resources, politics and influence (HPI). These components must work together to achieve the "optimal" value of the system. Similar relationships exist for risk and opportunity, which can be thought of as deviances from the expected value of the system. Thus, there is a higher-level structure operating between value, opportunity and risk in a system. The degree to which a system achieves maximum value, minimizes risk and maximizes opportunity is limited by the constraints on, and resources available to, the system's decision makers.

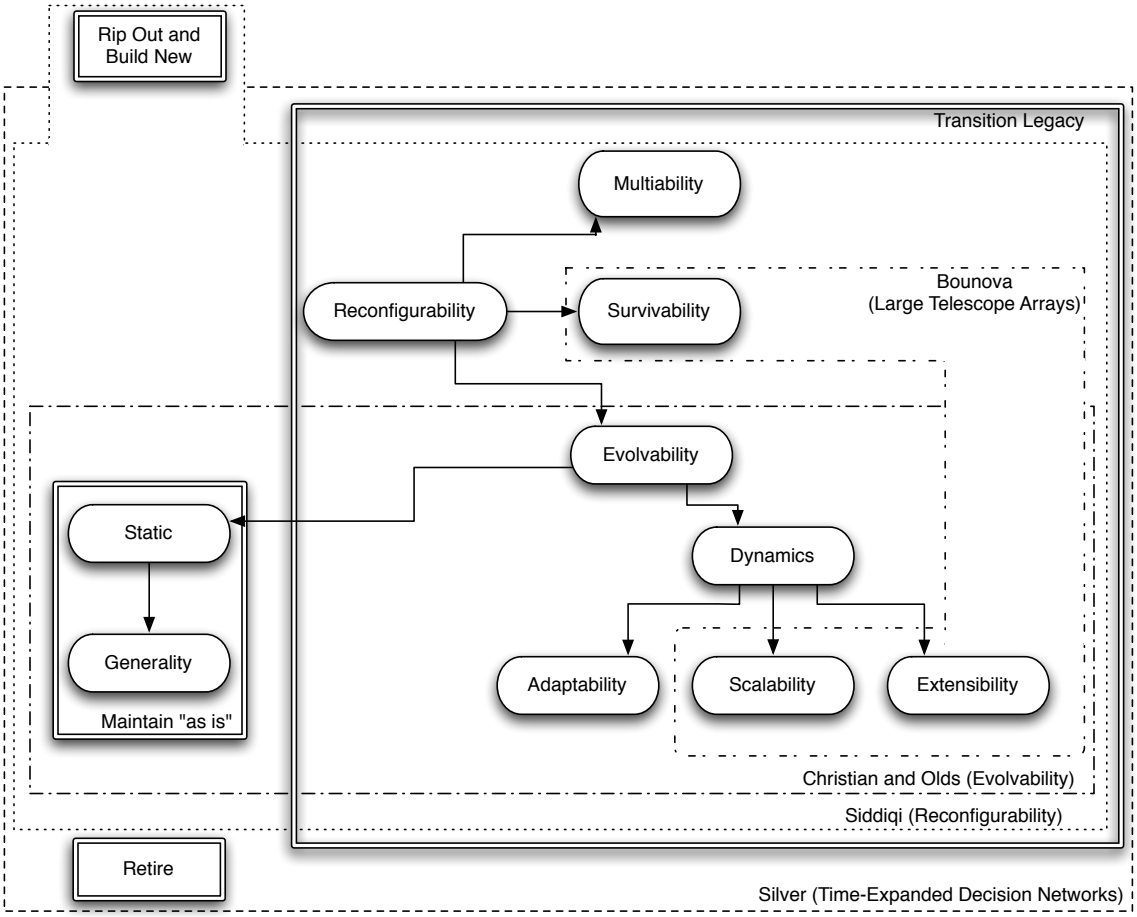


Figure 1.3: Relationships between key evolution research areas separated by the decision options they enable (e.g., transition (modify) legacy, rip out and build new, maintain "as is" or retire). The contributions of four specific pieces of work are overlayed onto this map.

1.3.4 The Evolution of Complex Systems

This section previews the relevant literature on the evolution of complex systems.

Figure 1.3 depicts the relationships among numerous key research areas (e.g., reconfigurability, multiability, survivability, evolvability, generality, adaptability, scalability and extensibility) and overlays the contributions of four specific pieces of work that will be discussed in this section. The research areas are separated by the decision options they enable: transition (modify) the legacy system, rip out the current system and build a preferred system, maintain the current system "as is" or retire the current system.

According to Siddiqi [25], reconfigurability enables multi-ability, survivability and evolvability. Multi-ability refers to the ability of a system to "perform multiple, distinctly different functions at differ-

ent times." Survivability refers to the system's ability to remain functional when components fail, though not necessarily at the desired level of performance. Multi-ability is categorized as a transition legacy decision option since the property allows the system to change to accommodate evolving functional requirements. Similarly, the survivability property enables the system to change to accommodate failures and degradation.

Evolvability, defined by Siddiqi[25] as the property of a system that "allows the system to change easily over time by removing, substituting and adding new elements and functions" (i.e., expansion, upgrades, etc.), is clearly a transition legacy decision option. Evolvability has been defined in a number of different contexts. Of importance is the breakdown of evolvability in Christian and Olds,[26] where evolvability is defined as "the capacity of a system to adapt to changing requirements throughout its life cycle without compromising the integrity of the system." Here it is noted that evolvability has both static and dynamic properties, while Siddiqi acknowledges only the dynamic aspect. The static property is referred to as **generality** and represents the case in which the system already meets the "evolved" requirements and therefore does not need to be changed. In the language of this thesis, **generality** is equivalent to the maintain as is decision option.

The dynamic properties of evolvability are identified by Christian and Olds[26] as adaptability, scalability and extensibility. These are defined as "rearranging or duplicating existing system components within the current architecture," "increasing the size of architectural components to accommodate increased loads," and "adding new components that were not part of the original architecture," respectively.

Based on these definitions, Bounova's[27] work on Large Telescope Arrays can be encapsulated by scalability and extensibility. Here, evolution is defined as "changing fundamental form and/or function", while extensibility is "preserving the nature of the elements of the system and their function, while increasing their number or size." The concepts of size and number[25] are likewise joined in the increased/added and decreased/removed reconfiguration processes. Number and size have also been separated into scalability and extensibility.[26]

Bounova models evolution via forward and backward staging. The forward staging technique optimizes the initial design and augments the second stage given the initial design of the legacy. The backward staging technique optimizes the final architecture, and the initial architecture is an optimally reduced version. The forward staging technique mimics the impact of legacy, while the backward staging technique behaves like a planned evolution. Interestingly, it was found that back-

ward staging was not the best strategy for the design of staged telescope arrays. Bounova also examines the impact of failures on robustness in large telescope arrays, encompassing survivability, in Figure 1.3.

Silver[15] accounts for the retirement option by adding a retirement sink node at the end of the Time-Expanded decision network. Although the article does not account for retirement costs, the method makes it easy to incorporate them in future research. There is, however, no explicit rip-out and build-new option. The methodology is demonstrated using Heavy Lift Launch Vehicles which are capable of launching more than 30,000 pounds to low earth orbit. In a sense, the methodology models the switching costs for product platforms or evaluation of whether a set of architectural options can be considered product platforms, but does not extend easily to capture system-of-systems such as distributed networks. Furthermore, the method is geared toward designing more flexible complex systems by identifying ways to reduce switching costs.

Siddiqi[25] further proposes an evolvability metric that enables a comparison of the results of reconfiguring the existing system or building and deploying a new system to meet the changed requirements. This metric can be extended to examine the desirability of the rip-out and build-new case by making sure the proper costs and benefits are accounted for appropriately. The metric is useful when it is clear from the outcome that either reconfiguration is most desirable (metric > 1) or ripping out and building new is most desirable (metric < 0). When the metric falls between these two values, it is unclear which option is more desirable.

Evolution Taxonomy

There seem to be a number of ways to describe the evolution of complex systems. This taxonomy of evolution is described below.

The evolution can be planned or unplanned.[25] For example, a technology roadmap[28] that allows for resources or technology to become available versus leaving the future configurations undefined until uncertainty resolves.

Transitions between architectural configurations can be reversible[25] or irreversible. Irreversible transitions reduce system flexibility and increase operational risk.

In a strictly theoretical sense, system evolution can either be path-dependent or path-independent. In nearly all (if not all) technical systems, the evolution is path-dependent. The changes made are

often irreversible and exhibit differences in both benefits gained and resources expended to make the transition and to operate the new system. It is very unusual in real, complex technical systems for the path taken to not matter to the final outcome. Path-independence may apply to certain types of uncertainty that drive the evolution of the technical system (e.g., evolution of financial forecasts).

The transitions can be intra-element (reconfiguration/modification of constituent elements) or inter-element (reconfiguration of the system of elements, scaling, etc).[25] Hybrid systems could be constructed in which transitions can be both intra-element and inter-element, such as adding a constellation of satellites constructed with different hardware.

Unconstrained evolution decisions are once again an idealization of reality. For example, one could design for a green-field system and choose the most evolvable system based on the methods outlined in this thesis. In the strictest sense, even this example is constrained since there are constraints spanning the realms of cost, policy, and etc. Another example might be the rip-out-and-build-new option, though even here there are constraints (similar to the green-field) but also constraints in terms of legacy processes, human resources, and etc. In other words, the old way of doing business but on a new system. It should suffice to loosely equate unconstrained evolution with green-field systems (systems in which the system is ripped out and built new). Constrained evolution refers to the legacy system case, in which there is an existing system and the changes are therefore constrained according to the needs and structure of the legacy.

Types of Change

From the above discussion, it is clear that there are a number of ways systems can be changed. The types of changes described are:

- Form
- Function
- Number (i.e., added, removed)
- Size (i.e, increased, decreased)
- Transform/re-arrange/re-distribute

For systems using a modular design, Baldwin[29] argues that there are six "admissible" evolution operators (note: list quoted directly):

- **Splitting** an interconnected task structure into modules;
- **Substituting** one version of a module (or set of modules) for another;
- **Augmenting** the design by adding a module with new functions at a particular interface;
- **Excluding** a module from a system;
- **Inverting** a recurring design element into an architectural module;
- **Porting** a module to a new system.

Baldwin's set overlaps with the taxonomy identified using the four pieces of work described in the previous section. It extends the taxonomy by acknowledging changes to task structures and revisiting the design to potentially identify options that were previously unseen.

1.4 Toward Strategic Evolution

This section will develop the concepts that will be used to evalute the history of the Deep Space Network in Chapter 2. The concepts bridge the gap between the background described in the previous section and the theory introduced in Chapter 3 and will be confirmed by applying them to the issue of transitioning from IPv4 to IPv6.

On the basis of the discussion in the previous section, this thesis proposes the following definition for the strategic evolution of systems:

> *The intelligent employment of strategy to achieve the long-term advantage of a system in operation, by way of maintenance of the system as-is; modifications to an existing component(s), interface(s), protocol(s), structural element(s) (legacy); retirement; or by the wholesale replacement of said system (rip-out and build new).*

The application of multi-attribute theory to building rehabilitation versus redevelopment decisions provides a real-world example of the difficulty of choosing between rehabilitation (i.e., transition to "new" system) or ripping out and building new.[30] The high-level decision options available to a decision maker using strategic evolution are diagrammed in Figure 1.4 on page 54. Changes occurring within an architecture family are separated from changes made between architecture families. The legacy inertia increases substantially when a decision maker considers moving between architecture families. When changes occur within an architecture family, the options are further separated by whether the changes occur within an element (intra) or between elements (inter).

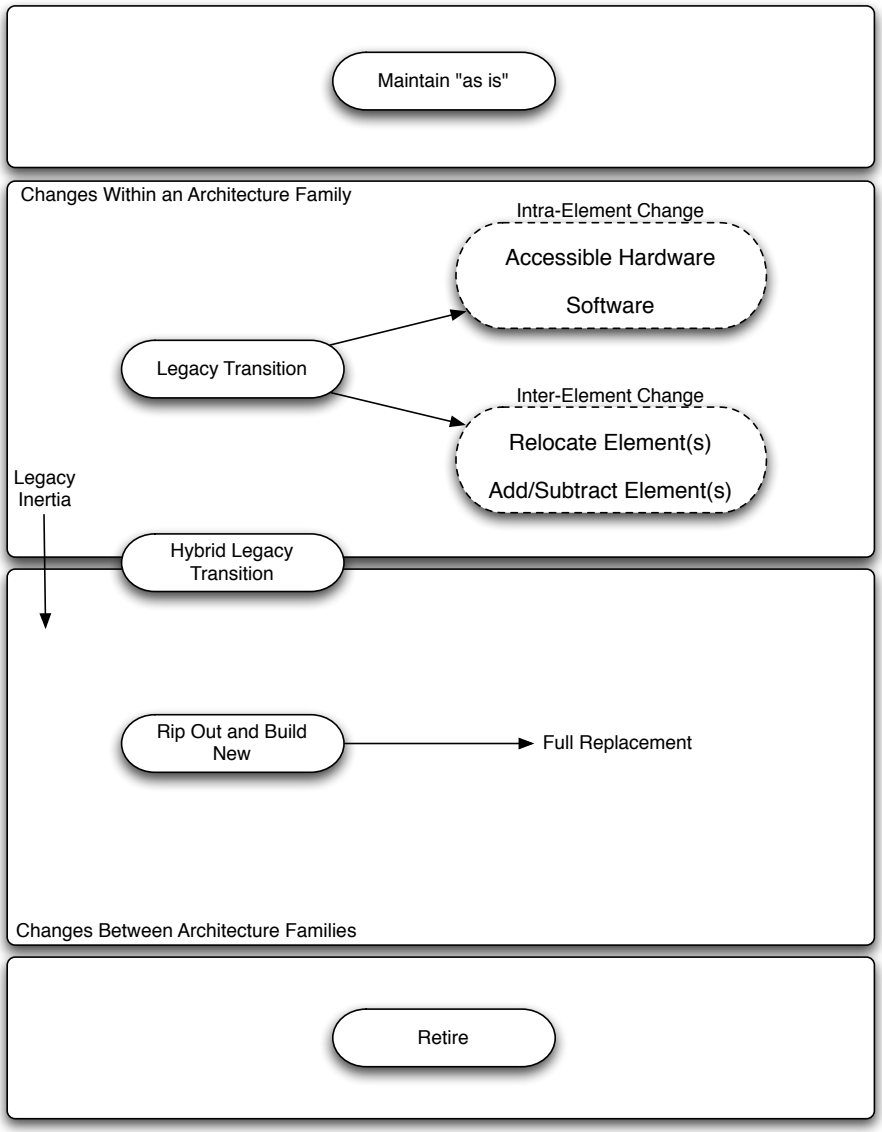


Figure 1.4: Diagram relating the various high-level options available to decision makers. Changes occurring within an architecture family are separated from changes made between architecture families. The legacy inertia increases substantially when a decision maker considers moving between architecture families. When changes occur within an architecture family, the options are further separated by whether the changes occur within an element (intra) or between elements (inter).

1.4.1 The Process Models

Here the process models involved in making system evolution decisions are proposed.

Any methodology for the strategic evolution of systems must occur within the context of the environment and legacy system, the external and internal drivers for change, and the organization processes that evaluate these events, determine and evaluate the options and strategies, make the

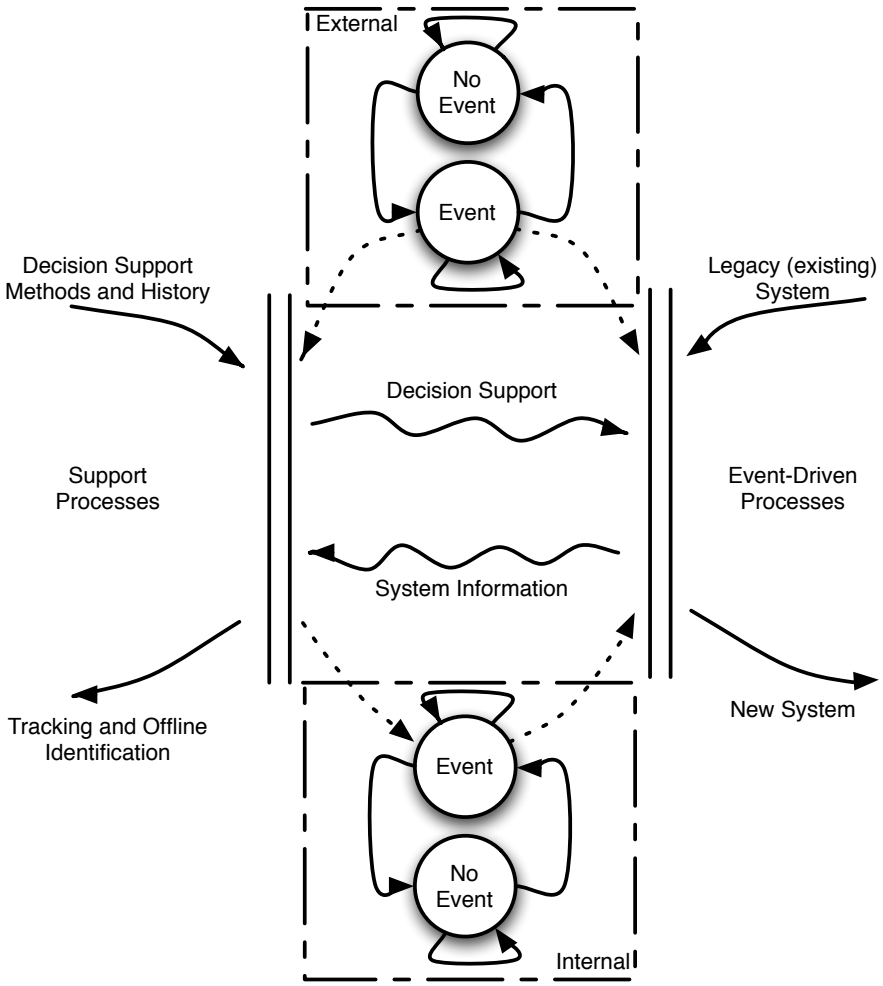


Figure 1.5: A process model that generically describes the high-level interaction between the key elements of system evolution: events that trigger the need to change (internal and external), the support process that manages the models and detailed analyses, and the decision process that evaluates the options — based on the information provided by the support process — and makes the decisions.

decisions, and implement, operate and monitor the consequences of the decision.

For this reason, this thesis proposes a high-level process model that generically describes the high-level interaction between the key elements: events that trigger the need to change (internal and external), the support process that manages the models and detailed analyses, and the decision process that evaluates the options based on the information provided by the support process and makes the decisions. The formal structure of the organization can take many forms, but the generic high-level view can be simplified to the basic components in Figure 1.5.

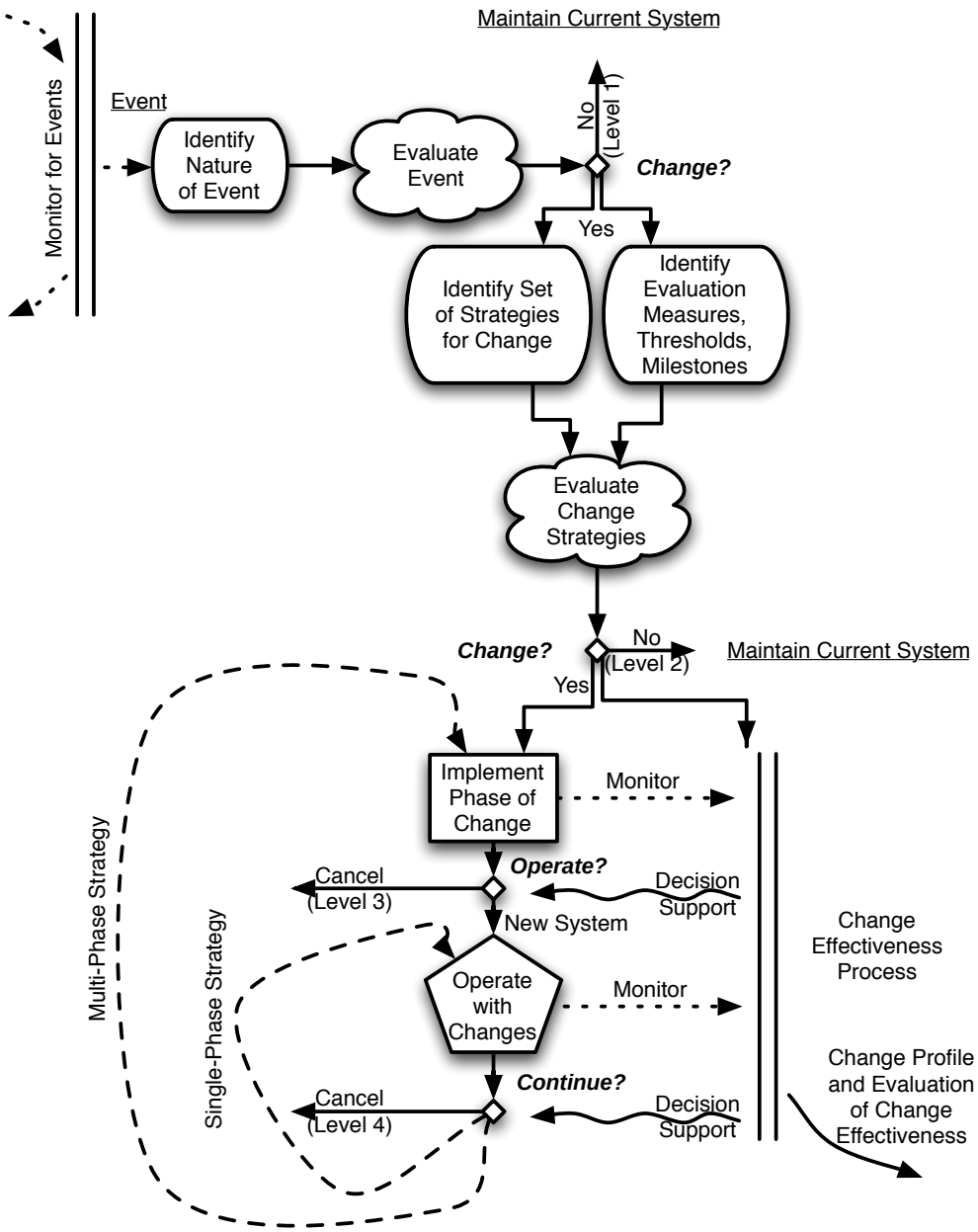


Figure 1.6: The event-driven process model (the decision-maker process) describing the flow of decisions from an internal or external event, to the support and decision processes determining a solution to the implementation and operation phase with monitoring.

Example external events:

- Decision parameter change (e.g., processor speed, component prices)
- Launch of a competitor
- Service/product demand change

Example internal events:

- Decision parameter change (e.g., spending caps)
- Component degradation
- Component failure

Events can be categorized as either opportunities (a situation or condition favorable for attainment of a goal) or threats (an indication or warning that the goals may not be met).

The event-driven process in Figure 1.6 describes the flow of decisions from an internal or external event, to the support and decision processes that determine a solution to the implementation and operation phase with monitoring. In this sense, the event-driven process describes system evolution.

The following section demonstrates these processes with a real-world example.

1.4.2 Example: Transitioning from IPv4 to IPv6

This section will highlight the key ideas presented in the previous discussion as applied to businesses deciding whether or not to make the transition from IPv4 to IPv6.

IPv4 has been used for over 30 years and IPv6 is now maturing. Although the U.S. Department of Defense has internally mandated the transition with core capability by 2008 for their systems and the U.S. government has since done the same for its agencies, businesses are reluctant to make the switch. Although the proclaimed reason for the transition has been the threat of depletion of address spaces, businesses perceive that there are much more cost-effective ways to minimize this effect via the use of Network Address Translation (NAT) and/or proxies to hide the network and access private IPv4 address space. CISCO and Juniper have been adding IPv6 functionality to their routers since 2001 [31] and Microsoft is now supporting IPV6 and Next Generation TCP/IP stack in their newer and emerging Windows products.[32] With the ability to run IPv4 over IPv6, in combination with NAT and proxy address space fixes, it may be some time before businesses decide to make the change. This example will show how the strategic evolution process models may be applied to businesses weighing the IP transition.

The high-level process model identifies two critical sub-processes: the event-driven processes (that which is typically associated with change and is conducted by the decision makers) and the support

processes that enable strategic decisions to be made. Figures depicting the high-level model and the event-driven processes are on page 55 and 56, respectively.

Various reports have emerged that IPv4 address spaces are projected to run out between 2010 and 2011.[33,34] If these reports can be considered an external event, then the strategic evolution process models suggest businesses should re-evaluate their stand on transitioning.

The support processes for many businesses seemingly indicate that the use of NAT and/or proxies is sufficient reason not to transition. The decision support that such support processes provide must be based on sound reasoning that itself is based on accurate and complete information. This rationale is telling, because while depletion of address space was the main motivator behind the development of IPv6, it is not the only change to the protocol. In fact, there are several key functionalities that have been introduced with IPv6 that can not be easily replicated (if at all) in IPv4. Many enhancements were made on the basis of experience dealing with the patched and often-labored functionalities in IPv4. In other words, the main motivator for IPv6 for many companies may not be the threat of address space depletion, but the opportunity that exists for the creation of so-called "killer" applications and improved competition due to Quality of Service (QoS) enhancements.

The first step in the event-driven process is to identify the nature of the event, which in this case is the expected depletion of address space by 2011. The second step is the evaluation of this event, a process that is strongly influenced by the support processes for the business. Currently, it appears that the business decision makers believe that transitioning to IPv6 will unnecessarily increase the address space at a cost with no additional benefit. Based on this information, the decision makers will likely decide to maintain the current system because they believe they will lose money if the transition is made. In other words, they would spend money that could be invested elsewhere in the business for no perceivable benefit.

Now, let it be assumed that complete and accurate information is provided to the decision makers during the evaluation phase.

Identifying the Nature of the Event

The Strategic Evolution of Systems methodology seeks to help decision makers make the most strategic decision possible given the information available. Success of the methodology will depend on the completeness and accuracy of the information.

It is reasonable to assume that the motivation to transition to IPv6 should not be based solely on threats to a business, but rather it must also include consideration of the opportunities that transitioning enables. This section briefly evaluates the threats and opportunities the transition from IPv4 to IPv6 presents. This outlines the information and evaluation that should be done by the processes that support the decision making (one of many functions the support processes should perform "off-line".

Threats

Here are two possibilities for threats (indications or warnings that goals may not be met) deserving of response:

First, suppose IPv6 penetrates the business world over time and becomes "the" standard for one or more segments of the business community. At this point, a company in one of these segments risks losing their competitiveness unless they also transition to IPv6.

Second, suppose 2011 is right around the corner and the world really is running out of address spaces. Further suppose the patches to IPv4 are insufficient to support the application requirements of the time and there are no immediate fixes to the problem. This motivator is technically the same as the current warnings, except the problem is evident and immediate action is required.

The first threat is more realistic based on historical evidence and the fact that Commercial Off The Shelf (COTS) hardware and software supporting IPv6 is becoming increasingly standard. It is tempting to term this the "peer-pressure migration."

Opportunities

A detailed examination of the differences between IPv4 and IPv6 seems to provide more opportunities than most articles on the subject give it credit.[31,35]

Here are a few of the most critical functionalities that IPv6 enables:

- IPv6 provides support for QoS so that routers can prioritize packet distribution.[31,36] For example, streaming data can actually stream with very little, if any, interruption.
- IPv6 supports an entirely new service functionality using the "anycast" routing option. This routing option enables users to send messages to "the easiest-to-reach member" of a chosen group.[36]

- Simple, straightforward definition of scope for "multicast" broadcast routing.[36]
- The datagram headers embed options in a very clever way, lending the headers to future applications, control options and etc., that may not be conceived of yet. When the options are not needed, the header space is not used, so the packet is as small as possible, lending itself to greater efficiency and higher data throughput.[36]
- Hierarchical addressing (and enhanced auto-reconfiguration functions), which enables useful things like the reflection of router hierarchy on Internet Protocol (IP) addressing (imagine, rapid aggregation of routes and packets for more efficient routing and being able to look at an IP address and immediately have useful information about what and where it is).[36]
- Finally, the huge address space can enable extremely useful features such as directly mapping Media Access Control (MAC) addresses to IP addresses and real support for end-to-end transport, security and etc.[36]

Many of these critical functionalities are likely ignored because the benefits are "intangible." The main "intangible" benefit of a transition to IPv6 is that it enables and supports — by providing flexibility with addressing, un-designated header options and "anycast" routing — new services and functionalities, some of which may not have been thought of yet. One cannot put a value on that which one does not know anything about.

There are several critical "tangible" benefits of transitioning to IPv6 that are apparent by looking at this partial list of new and enhanced functionalities. The first is increased efficiency. More efficient routing and overhead means less congestion, better round-trip delay, and more capacity available on the network for real data. This added efficiency not only improves QoS, but it also increases data throughput for a given cost outlay. QoS support enables guaranteed QoS to customers, which will improve customer retention if used wisely. The improvement in auto-reconfiguration, addressing structure, mapping capabilities and routing means companies can spend fewer resources on establishing and securing their networks since the new functionalities require fewer brute-force methods such as the need for static routing in many IPv4 networks. Additional "tangible" benefits exist, but the ones identified here are sufficient for this example.

These "tangible" and "intangible" benefits must be included in the decision mechanisms in order to provide and accurate and holistic view of the consequences of transition or not.

Event Evaluation

The previous section identified threats and opportunites that motivate a change evaluation. The next step is to use this information to address the question of whether (and when) to transition from IPv4 to IPv6.

When should a business choose to transition to IPv6? Some businesses rely heavily on IP-based data communications because they are data communication service providers. Others merely use IP-based communications to network their employees within the company and to the outside world. Every business relies on IP-based communications to some extent, and when they should transition strategically depends heavily on the extent of that reliance and how it impacts their bottom line. Some businesses might, with the right IP-based services, transition from low-reliance to high-reliance. If the perceived benefits outweigh the expected costs and associated risks of transitioning, then the business should transition. The difficulty comes in weighing these pros and cons, especially when attempting to project the consequences beyond the immediate future. Consider a few examples of these trades.

Suppose the intangible benefits provide future options for business expansion that are not attainable from the current architecture. The effect of not transitioning is to limit that business-expansion option space. The company is now exposed to the risk of losing market share or even losing out on being the first out with a new "killer" application or a patent of ideas and the ensuing royalties. Transitioning to IPv6 effectively embeds options for the future during system operation. This "intangible" benefit is key for businesses that rely heavily on providing IP-based data communication services. These businesses would strategically transition to IPv6 for this reason, notwithstanding the "tangible" benefits of QoS control, efficiency, and etc. Once one of these companies successfully transitions, the rest will necessarily have to follow to stay competitive and to continue to hold on to their market share.

Businesses that rely moderately on IP-based communications might have a tougher time deciding. These companies will need to identify whether the company can leverage the benefits of IPv6 to position themselves strategically for developing, adapting, and/or implementing applications, services and products in a timely fashion. Businesses in this category may decide to transition because it enables them to expand their business into other areas. Others may choose not to make the transition because they wouldn't be able to leverage enough of the features to make it worth their while (i.e.,

the difference in expected revenue coming in over some reasonable time frame may not cover the investment outlay to transition). The business will need to be able to answer the following questions:

1. Does this change increase the value (benefit at cost) of the overall system over the time horizon or does it decrease it?
2. Does this change improve or hinder our ability to respond to future external change?
3. Does this change improve or hinder our ability to expand our system/processes/product (internal change)?

When should a business choose to NOT transition to IPv6? Businesses should not transition to IPv6 if the cost and risk of transitioning outweighs the perceived benefits. This situation is likely to occur in small businesses that do not provide IP-based communication services to customers, and with few networked employees. For these businesses, the trade will often hinge on the resources spent on establishing and securing the network. The cost of transitioning may not be worth the switch since the benefits are minimal if not negligible (the exception might be a business in which security of the network is a necessity, although there are some arguments against the true security of IPv6 relative to IPv4).[35] If the business already uses software and hardware that supports IPv6 and the Information Technology (IT) team is already familiar with it, then the costs to transition are minimal, but the necessary benefits may not exist to motivate a change.

If a trade favors staying with IPv4, then the company should maintain the current system unless market competitiveness forces them to transition down the line. It is interesting to note that this is an example of a network effect; without a sufficient number of users adopting IPv6, the incentives to switch are low.

Strategically Transitioning from IPv4 to IPv6

Once the decision has been made to transition from IPv4 to IPv6, the company must figure out the best strategy for doing so. The strategic-evolution-process models indicate that the company must identify a set of strategies for change as well as evaluation measures, thresholds, milestones and any other quality attribute that is required for determining the effectiveness and benefit of the transition.

Identify Strategies for Change According to Pujolle, there are three main types of mechanisms for transitioning IPv4 to IPv6.[37]

1. **Double pile:** routers and terminals support both IPv4 and IPv6 protocols; has same addressing "problem" as all machines must have a public IPv4 address.
2. **Encapsulation:** IPv6 packets encapsulated in IPv4 packets.
3. **Translation:** translate between the two versions so IPv6 subnetworks can utilize the IPv4 infrastructure.

These mechanisms enable the co-existence of IPv6 networks with IPv4 to smooth the transition for as long as it takes for the entire network to become IPv6 compliant. For some businesses, it is as simple as saying "these are the strategies." For others, there might be hybrid strategies if the architecture of the business network allows it, which may also give rise to staged deployment strategies if that seems feasible to the decision makers.

Identify Evaluation Measures, Thresholds, Milestones Example evaluation measures are the cost of the transition and the interruption to the business during the transition.

The cost of transitioning is impacted by several factors: hardware upgrades (e.g., routers), software upgrades (i.e., licensing Windows XP instead of remaining with Windows 98), and human resource training for those employees impacted by the change (e.g., the IT department).

The cost of the hardware upgrades will be a function of how many components will need to be upgraded to end up with the desired end architecture. A similar point can be made about the software due to licensing requirements. In some businesses, the software issue is more complicated than simply upgrading to Windows XP, and in these cases the costs could be significant if software customization is required.

Evaluate Change Strategies The previous sections covered the information required to make a decision concerning the mechanism for transitioning the business from IPv4 to IPv6. The next step is to make that decision. The business has the right to refuse change at this point and to continue to maintain the current system. This result would occur if the cost to transition ends up being prohibitive for that particular business.

The key idea behind evaluating the change strategies is to find the transition solution (if one exists) that meets all of the requirements for the transition. These will be business specific and would need to be based on a thorough review of the structure and needs of the business.

The exact mechanism for transition — or co-existing with IPv4 networks — a company chooses will depend on the architecture of the business. A business providing IP-based services will naturally need to maintain some form of co-existence everywhere for customers who insist on staying on IPv4. Businesses with fairly self-contained networks can move fully to IPv6 internally and simply translate at their company gateways. In some businesses, the transition strategy might be a staged deployment of hardware and software for only those segments of the business that require it at any given time.

Implementation and Operation Stages If the decision to transition does occur, then the business must monitor the implementation in order to minimize disruptions to the current system and to ensure that the change is meeting the desired effectiveness. Finally, the business must decide whether to actually operate with the change. Even if the business transitions the hardware to IPv6 compliance, it does not mean that the business must continue with the upgrade of the software if they decide it is not feasible at that point in time. The business can of course come back to re-evaluate that decision. Furthermore, even if the business has been fully operating on IPv6, the framework provides the business the opportunity to evaluate whether IPv6 is really as effective as IPv4 for their particular requirements. The business may decide to switch back to IPv4 (possible since IPv6 is backwards compatible) if, for example, using IPv6 interferes with their ability to interact with other critical businesses that are not using IPv6.

Conclusion

This example applies the strategic evolution process models to help businesses weigh when (if at all) to transition to IPv6 from IPv4 and how they should accomplish the transition. The discussion motivates the usefulness of the process models and the necessity of having tools to quantify the operational exposure to making changes.

1.5 Thesis Contributions

The four main contributions of this research are:

1. Establishment of a new framework for the evolution of complex technological systems that naturally incorporates legacy.

2. Articulation of two important principles of system evolution based on empirical observation of NASAs Deep Space Network (DSN) and the transition from IPv4 to IPv6.
3. Preliminary formulation of a position entropy metric, derived from information entropy, to describe the current position and evaluate the desirability of potential future transitions of system configurations.
4. Application and implementation of the framework to two forward-looking case studies: a commercial satellite communication system, and the staged deployment of communication infrastructure to support manned lunar exploration.

Secondary contributions include:

- Definition of the concept of option lock-out.
- Proposition of the process models for strategic evolution.
- Identification of an energy analogy to model static architecture desirability.
- Definition and conceptualization of the "position" of an architecture relative to its alternatives, which enables identification of the most advantageous immediate transition.
- Establishment of a method to visualize the structure of transition options using supernodes, or the subset of architectural configurations that are fully reversibly connected.
- Development of visualization tools that depict the near-term and long-term exposure to uncertainty if a given transition option is exercised. These tools include the Strategic Advantage Map and the Iceberg Exposure Graph, which maps the exposure of net present value for each accessible transition option relative to a neutral no-gain-no-loss line, resulting in a graph resembling an iceberg.
- Joint evaluation of inter-reconfiguration and intra-reconfiguration.
- Extension of a market model to evaluate attainable business segments in global satellite communications by integrating estimates of the global distribution of market demand, observed traffic statistics and calculations of the resulting steady-state network performance.
- Modeling of QoS applications, service and business regions.
- Extension of the Architecture Decision Graph methodology to the evolution of complex systems. The methodology was originally intended to prioritize and intelligently guide design and development efforts.

1.6 Thesis Overview

Chapter 2: A historical motivation for the Strategic Evolution of Systems, examining the history of the Deep Space Network in two parts. The first part describes the early history of the network, including identification of examples of items discussed in the introductory chapter and an analysis of insights about what went right and what went wrong. The second part evaluates the evolution of the full history of the network on four levels: organizational change, mission evolution, physical architecture composition evolution, and technology improvement and infusion.

Chapter 3: A detailed description of the philosophy, principles, theory, framework and implications of the Strategic Evolution of Systems. The chapter including the derivation of a new metric for evaluating the long-term position of an architecture.

Chapter 4: A high-fidelity case study of a commercial satellite communication system undergoing orbital and software reconfiguration. The case study demonstrates the application of the Strategic Evolution of Systems framework for a monetary-based complex technological non-market traded system.

Chapter 5: A low-fidelity case study intended to illustrate how to qualitatively define the energy for a complex technological non-market traded system that is **not** defined in terms of monetary units, nor does it have the option to maintain the system "as is." The case study evaluates the position of a communication infrastructure that is deployed in stages to support the manned exploration of the lunar surface, measured relative to a benchmark network that is deployed upfront.

Chapter 6: Conclusions: briefly summarizes the objectives, approach, major results and contributions of this thesis, evaluates its shortcomings, advantages and potential applications, and discusses potential avenues for future work and other recommendations for extending the research and case studies.

Part I

DEVELOPMENT OF THE STRATEGIC EVOLUTION OF SYSTEMS

Chapter 2

The Evolution of the Deep Space Network

This chapter provides historical motivation for the strategic evolution of systems by examining the Deep Space Network (DSN) in detail. The DSN is a collection of antenna facilities that support space assets during missions, enabling transmission of commands and reception of scientific and technical data. The objectives of this historical study are:

- To understand the influence of various decision factors (known, unknown, political and technical) on the design, evolution and corresponding success (or failure) of systems in operation.
- To gain insight into the types, mechanisms and consequences of changes during operation by studying organizational change, mission evolution, physical architecture composition evolution, and technology improvement and infusion.
- To achieve confirmation of ideas expressed in the introductory chapter.
- To study a historical example in order to develop principles and philosophies key to successful strategic evolution of complex systems.

The first section discusses the events leading up to the creation of the DSN, providing insight into the factors shaping the initial design. It is written as a historical narrative with commentary related to building the Strategic Evolution of Systems theory included as footnotes. Process models based on the historical narrative are included in Appendix B.1. The second section analyzes the history of the DSN from the perspective of its evolution, identifying relevant patterns and mapping out

the reasoning for changes and their subsequent consequences. In addition, specific examples of evolution that provide significant insights and confirmation of the ideas presented in this thesis are described in detail in Appendix B.4.

2.1 The Creation of the Deep Space Network

In 1950, a group of geophysicists met in Silver Spring, Maryland, with Sydney Chapman, a visiting English scientist. Ostensibly, they gathered to discuss Chapman's recent theories concerning Earth's atmosphere. The exchange shifted – unsurprisingly – to geophysics, and methods to measure and observe the Earth and upper atmosphere concurrently. Many new revolutionary scientific tools had become available, including rockets.[1] The head of the new Brookhaven National Laboratory, Lloyd Berkner, suggested that these advances should be leveraged to study Earth's systems[2] in a manner similar to the International Polar Years of 1882 and 1932.[38,39]

2.1.1 The International Geophysical Year: Setting the Kindling

Berkner and Chapman collaborated to further develop the idea for a new, extended International Polar Year. The original International Polar Year had focused international scientific cooperation on studying conditions in the polar regions. The new proposal would encompass the entire earth.[3] The scientists presented their matured idea to the International Council of Scientific Unions.[40,39]

The International Council of Scientific Unions endorsed the plan in 1952, renaming the project the International Geophysical Year (IGY) to capture the expansion in scope.[4]

Establishing the International Geophysical Year

A Special Committee for the IGY, also known as Comite Speciale de l'Annee Geophysique Internationale (CSAGI), was appointed to oversee the necessary preparations and to perform the role of governing all IGY activities. The Council scheduled the IGY for 18-months between July 1, 1957, and December 31, 1958, to take advantage of several eclipses and an expected period of maximum sunspot

[1] Paradigm shift in technology produces new opportunity.
[2] Primary opportunity identified.
[3] Primary opportunity system scope defined (space).
[4] External event for Department of Defense DoD.

activity.[5] [40,38,39]

During the seven years leading up to the IGY, scientists from 46 countries planned events and prepared projects to support the "series of coordinated observations of various geophysical phenomenon." By the end of 1958, another 21 countries had joined the effort. [38,40]

March 1953 saw the National Academies appoint the U.S. National Committee (USNC) to oversee the American contribution to the IGY. The USNC formed technical panels to research and develop experiments on topics ranging from oceanography to seismology, gravity to solar activity and cosmic rays, glaciology to geomagnetism, rocketry to longitude and latitude determination.[6] Most importantly, the USNC wanted the United States to launch an artificial satellite into Earth's orbit.[7] [40]

The Soviet Response

CSAGI worked hard to keep the IGY focused on science, preventing any injection of international politics. The goal was international collaboration, with the CSAGI operating as a "nonnationalistic, apolitical" body.[8] [41,38]

However, the United States and the Soviet Union had been involved in the Cold War[9] since the mid 1940s. The Cold War was an artifact of differences surrounding post-World War II (WWII) reconstruction, but the ideological clash between Bolshevism and capitalism continued. [42]

November 1953 brought a statement by A. N. Nesmeyanov of the Soviet Academy of Sciences that "satellite launchings and moon shots were already feasible." The United States had reason to believe the Russians were capable of demonstrating the feasibility of spaceflight as Tsiolkovskiy's work on mathematically formulating modern astronautics was well known. Moscow Radio upped the ante in March 1954 when it called for "Soviet youth to prepare for space exploration." A month later, a disclosure by the Air Club in Moscow revealed that "studies in interplanetary flight were beginning."[10] [39]

[5] Secondary opportunity identified and scope defined (time).
[6] Support process: development of knowledge.
[7] Decision to "compete" — Project goal.
[8] Politics.
[9] International politics.
[10] Political positioning; shaping factor for DoD event evaluation.

The Stakes are Set

Discussions leading up to the CSAGI meeting in Rome in the late summer of 1954 made it clear that "political and psychological prestige" would befall the "nation that first launched a man-made satellite."[11] [39]

The Union of Soviet Socialist Republics (U.S.S.R.) remained aloof during the CSAGI meeting; the Soviet representatives listened, but did not officially join the IGY nor speak up during the discussion. [39]

The Adoption of the Satellite Launch Resolution

Following the CSAGI discussions in Rome, the Council formally adopted a resolution on October 4, 1954, calling for the Earth's surface to be mapped during the International Geophysical Year using artificial satellites.[12] [39,43].

> *In view of the great importance of observations, during extended periods of time, of extra-terrestrial radiations and geophysical phenomena in the upper atmosphere, and in view of the advanced state of present rocket techniques, CSAGI recommends that thought be given to the launching of small satellite vehicles, to their scientific instrumentation, and to the new problems associated with satellite experiments, such as power supply, telemetering, and orientation of the vehicle.* [44]

The Department of Defense Solicits Proposals

Seemingly faced with mounting pressure by the Soviet Union,[13] the U.S. Department of Defense solicited proposals in 1955 for an Earth-orbiting satellite from its Army, Navy and Air Force components[14] to satisfy the United States contribution to the International Geophysical Year. [1]

The Program Proposals

The Assistant Secretary of Defense Donald A. Quarles invited the U.S. Air Force in January 1955 to submit a proposal for launching a satellite with an Atlas Intercontinental Ballistic Missiles (ICBM).

[11] Anticipated consequence.
[12] Project goal.
[13] Part of event evaluation; Shaping factor.
[14] Identify set of strategies.

In response, the Air Force suggested a combination of an Atlas A ICBM prototype with an Aerobee 150 second stage.[15] The Atlas program manager, Bernard Schriever, was not thrilled. The Atlas ICBM program was top-priority, and he opposed any diversion.[16] [41,45]

At the time the proposal request went out, the Jet Propulsion Laboratory (JPL) was on contract with the U.S. Army to develop guided missiles at the White Sands Missile Range. The contract gave JPL an opportunity to develop significant guidance and tracking expertise for large rockets and ICBMs.[17] The Jet Propulsion Laboratory and the Army responded to the Defense Department request by proposing the Orbiter satellite[18] in September 1954. [1,41]

The Project Orbiter proposal included concepts developed by Wernher von Braun and his rocket team. The architecture was fairly simple: a Redstone with the upper stages supported by clusters of Loki rockets. The reliance on existing hardware and facilities enabled Orbiter to claim the lowest cost and technical risk of any of the satellite proposals for the United States contribution to the IGY. [45]

Project Orbiter was a joint proposal from the U.S. Army Ordinance Corps and the Office of Naval Research, which was the parent organization of the U.S. Naval Research Laboratory (NRL). The NRL desired an all-Navy alternative to Project Orbiter and the Air Force's World Series vehicle,[19] and formally submitted their own scheme to the Advisory Group on Special Capabilities in July 1955. The NRL had already sold their design to the U.S. Academy of Sciences and the Head of the American IGY Committee.[20] [41,45]

The NRL proposed Vanguard, a mostly new vehicle derived from existing vehicles. The main stage was an elongated variant of an obsolete sounding rocket, fitted with an engine developed by a former U.S. Army rocket program. A derivative of an Aerobee sounding rocket fitted with an Aerojet engine served as the upper stage. [45]

[15] Support process: proposal response.
[16] Politics. Influence.
[17] Support process: knowledge and experience development.
[18] Support process: proposal response.
[19] Human nature.
[20] Influence.

The Department of Defense Chooses Vanguard

The Advisory Group on Special Capabilities were set to make their decision on the satellite proposals in September 1955. The selection itself would be heavily influenced by the political agenda[21] of the Eisenhower Administration. The Administration strongly desired the establishment of a legal precedent that would limit disputes with the Soviets over the launch of military missions.[22] Once the freedom of space was established, the United States could proceed with its planned espionage using the Corona satellites.[23] A 'civilian' satellite as the United States contribution to the IGY could provide such a precedent.[41,45]

As mentioned earlier, the ICBM program was a top-priority of the Administration, and any interruption was unacceptable. Thus, the proposal selected by the Advisory Group could not introduce even the slightest delay to the program.[24] [45] Eisenhower appointed the secret Stewart Committee in May 1955 to decide on "the best course of action." Each of the three branches of the military had two representatives on the Stewart Committee, with two additional members selected by appointment according to the Assistant Secretary of Defense Quarles. One of the appointed members was Richard Porter, one of the Vanguard designers. The other member was the same head of the American IGY Committee that the NRL had sold the Vanguard design, Joseph Kaplan.[45] The "best course of action" predictably became a political showdown:[25]

> *The Army and Navy representatives voted for their services' proposals. Kaplan was not about to see von Braun's "arrogant Nazis" get the job,*[26] *and voted together with Porter for the Navy proposal. The Air Force representatives were inclined to vote with the majority, and certainly did not want either the Army or the Germans to get the job. In August 1955 the Stewart Committee, having duly taken on the Pentagon's desire for the IGY effort not to affect either the Air Force Atlas ICBM or Army Jupiter ICBM programs, selected the Navy's Vanguard as the IGY satellite booster...Von Braun and his Army supervisor, General Medaris, fought this decision long and hard. But they were not only discouraged, but prohibited from launching a satellite.*[27] [45]

[21] Politics and influence.
[22] International politics.
[23] Triggering event to exercise option.
[24] Political constraint.
[25] Strategy evaluation. Politics, an inefficiency.
[26] Human nature.
[27] Politics and influence.

Thus, the Advisory Group on Special Capabilities dutifully voted for the Vanguard proposal seven-to-one on September 9, 1955. The U.S. Department of Defense immediately authorized the Secretary of the Navy to proceed with the NRL proposal.[28] Vanguard would be representing the United States during the IGY.[41,46]

Although Orbiter had been shelved by the Defense Department, the Army encouraged JPL to continue its low-level development over the next years.[29] [1]

2.1.2 The Launch of *Sputnik*: Igniting the Space Race

Within months of the Department of Defense acceptance of the NRL proposal, significant problems emerged that threatened the very existence of the Vanguard program. Unexpected personnel shortages in addition to costly and time-consuming redesigns left Vanguard in a precarious position.

The Development of Vanguard

A key premise of the Navy's original design was the availability of the Viking development team to aid in adapting the Viking sounding rocket to the needs of Vanguard's first stage.[30] However, the Viking team had mostly been reassigned to Titan, the Air Force's new ICBM program,[31] and were now completely unavailable.[32] [41]

Early on, new payload requirements forced a fundamental redesign of the Vanguard rockets. The selection of scientific instruments was mandated to the National Academy of Science's IGY panel.[33] The committee felt that the satellites should be used to study the density of the upper atmosphere, and eventually compromised for a 20-inch instrument sphere rather than a nose cone. The size increase of the Vanguard satellite meant that the proposed Aerobee second stage was too skinny,[34] and a larger diameter rocket would need to be designed by Aerojet-General. A third stage was the most risky component of the rocket system, requiring a technological advance in solid rocket motors.[35] To mitigate the risks, the Vanguard team contracted with two companies to design the

[28] Strategy and implementation decision.
[29] Politics and influence.
[30] Personnel critical path.
[31] Unwarranted assumptions and failure to adequately research critical path risks.
[32] Politics.
[33] Politics and influence; politics driving design driver decision.
[34] Unexpected change in design driver forces redesign.
[35] Technological critical path.

third stage in parallel.[36] [41]

While the rockets and payload were being developed, the construction of the infrastructure required to launch and support the satellite was encountering its own problems. The main launch facility for the Viking program in White Sands, New Mexico, was not a feasible launch site due to safety concerns for nearby communities.[37] A new facility for large ballistic missile launches was undergoing expansion in Cape Canaveral, Florida, and the Vanguard program worked out a build agreement for its own launch facility.[38] Political priority for military missile development again interceded, this time preventing the use of existing launch and tracking assets.[39] The Vanguard program contracted the development, construction and installation of the "Minitrack" tracking system to the Bendix Corporation. [41]

As the expected budget grew from the original $12 million estimate to almost $100 million, cancellation of the program was right around the corner. The Central Intelligence Agency (CIA), aware of the need to establish the freedom of space precedent, and sensing the enormous benefits of being the first in space, salvaged the program by infusing it with needed funds.[40] A live first-stage rocket test successfully launched on May 1, 1957, with the Minitrack tracking system coming online on October 1,1957. [45]

The Development of Orbiter

Meanwhile, the Jet Propulsion Lab spent the two years following the Defense Department selection of Vanguard building systems and infrastructures that would prove key to the eventual U.S. success in the space race.[41] Although JPL developed high-speed booster rocket upper stages and a critical phase-lock tracking receiver, most important was the development of Microlock, a high altitude test rocket tracking system. The flight unit carried a minimum-weight radio transmitter that communicated with the Microlock ground stations. Telemetry and positional data could be obtained from flight units up to a range of several thousand kilometers. [1]

While Vanguard verged on cancellation, the U.S. Army's General Medaris lobbied for Orbiter on the basis that von Braun and his team could launch a satellite with little notice at a fraction of the cost

[36] Redundant parallel design process; risk mitigation.
[37] Safety constraint.
[38] Leveraging legacy infrastructure with modifications.
[39] Politics.
[40] Politics and influence; propping up a failing program.
[41] Parallel support process.

of Vanguard.[42] Seven months before the successful first stage Vanguard rocket test, on September 20, 1956, the Army successfully launched the first three stages of von Braun's Redstone rocket. [45]

The Surprise Sputnik Launch

In the midst of the continued political battle between the U.S. Navy and the U.S. Army concerning the U.S. satellite program, the unthinkable happened.

Intelligence reports suggested to the Eisenhower Administration that the United States was further advanced with its rocket program than the Soviet Union.[43] Yet, only three months into the International Geophysical Year, on October 4, 1957, the Soviet Union successfully launched the first Sputnik Earth-orbiting satellite,[44] followed a month later by a larger version carrying a live dog. Even more startling, the Soviets had launched Sputnik using military ICBM, contrary to IGY insistence on non-military equipment.[45] Despite these concerns, the Soviets had established the freedom of space precedent,[46] clearing the way for the launch of the Corona espionage satellites. [45,38]

Although the public was aware of the IGY, it was unaware of plans to launch satellites into Earth's orbit. Americans panicked. Politicians and media commentators assumed a "missile gap" between the United States and the Soviet Union, leading to fear about what the Soviets planned to do next.[47] The United States had to respond. [38,45]

Following a status briefing, President Eisenhower released a statement on October 11, 1957, that Project Vanguard would soon launch its experimental Test Vehicle (TV)-3 test flight. The briefing had underscored the remote likelihood of a successful orbital insertion, but this did not prevent the press from opportunistically touting the test flight as a full-fledged satellite launch.[48] [41]

Vanguard successfully launched a second first-stage-only test 12 days later. Despite public complaints from von Braun that the United States could have been first in space had his project been chosen, Eisenhower continued to back Vanguard, apparently feeling von Braun's former Nazi status was politically untenable.[49] [45]

[42] Politics.
[43] Incorrect support process information.
[44] Triggering external event; threat.
[45] Aggression; shaping factor, making response more important.
[46] International politics.
[47] Shaping factors for the response.
[48] Human nature: misinformation and misinterpretation leading to unrealistic expectations.
[49] Politics.

The Reactivation of Orbiter and the Emergence of Explorer

Meanwhile, the U.S. Secretary of Defense Neil H. McElroy authorized the reactivation of the Orbiter program.[50] The Soviet demonstration of an operational, intercontinental ballistic missile capability ahead of the United States had sent up red flags at the Defense Department. Perhaps sensing problems within the Vanguard program, McElroy urged Orbiter to move toward a launch with all due haste.[51] [1]

The Orbiter program re-emerged as Explorer and prepared to launch in 90 days. The Army and Jet Propulsion Lab constructed Explorer using a Redstone rocket-based launch vehicle and a suitable satellite carrying a special science payload designed by Dr. James Van Allen to measure radiation.[52] The Microlock tracking stations were expanded to accommodate the mission,[53] adding stations at Cape Canaveral, Singapore, Nigeria and San Diego in California. Each of the stations in the United States had interferometric tracking antennas. The overseas stations provided telemetry and Doppler data via a single, helical antenna. Communications to the overseas sites sometimes meant trips on foot between the station and the nearest telegraph office.[54] [1]

The Public Failure of Vanguard

Then, in December 1957, the United States fell even further behind the Soviets. The first Vanguard launch failed miserably on live television when it exploded on the launch pad, leaving a stunned and embarrassed United States.[55] The space race became a desperate game of catch-up on the part of the U.S. government.[1]

> *On December 6, 1957 the first stage of TV-3 ignited and began to lift itself off. Only a meter (yard) off the pad and two seconds into the flight, the General Electric (GE) X-405 engine lost thrust. As if in slow motion, TV-3 settled back onto its launch pad, toppled over and exploded. The tiny grapefruit-sized test satellite rolled across the pad to safety and continued to transmit as TV-3 burned. The failure of what was supposed*

[50] Influence.
[51] Evaluate event, identify strategies, evaluate strategies; decision to reactivate orbiter.
[52] Opportunity to exercise payload option.
[53] Infrastructure modification; heterogeneous.
[54] Temporary solution given time constraints.
[55] Failure, crisis; bad public relations.

> *to be a test flight struck America to its core. Immediately Vanguard was dubbed "Kaputnik" and "Flopnik" by an unforgiving press. American leaders and the public lost faith in the program. While the next test flight, TV-3BU (for "TV-3 Back Up"), would fly within the next couple of months, the DoD and White House needed more options before this public relations disaster turned into a political and military one.*[56] *Von Braun's team would get their chance to launch a satellite.*[57] [41]

A Reversal of Fates: Explorer Launches Successfully

Within 60 days of the devastating loss of Vanguard, and a mere 84 days since its re-activation, Explorer I successfully launched[58] on January 31, 1958. History was made, and not just from the standpoint of Explorer I being the first U.S. Earth-orbiting satellite: the special instrument designed by Van Allen identified the existence of the Van Allen belt, a high-altitude region of radiation encircling the Earth.[59] [1]

On March 17, 1958, Vanguard 1 finally achieved orbit following a second launch failure. A congressional investigation was launched into the proposal selection process in an attempt to identify the party responsible for the apparent American loss of the space race.[60] [45]

The success of Explorer 1 served to place the Soviets in the role of catch-up. The Jet Propulsion Laboratory, under the "ambitious and far-sighted" direction of William H. Pickering, turned its attention toward deep space.[61] [1]

2.1.3 The Pioneer Program: from *Sputnik* to the Deep Space Network

Following the Soviet Union's success with Sputnik (see Section 2.1.2), Americans sought a revival of their international prestige. Many felt threatened by the Soviets preeminence. Corporations and institutions suggested ambitious space projects, including swiftly launching probes to the moon. More than 300 such proposals were sent to the Pentagon within six months of Sputnik. [47]

[56] Triggering event; now advantageous to expand option space.
[57] Time to shift strategy; support process brought "online".
[58] Good public relations.
[59] Scientific value.
[60] Politics and public relations.
[61] Vision.

Project Red Socks and the Pickering Proposal

Less than three weeks after the Sputnik launch but before the Vanguard failure, JPL proposed Project Red Socks.[62] The introduction underscored the necessity of an immediate substantial technological advance as Sputnik 1 "has had a tremendous impact on people everywhere" with a "significance which is both technical and political." [47]

JPL's director, William Pickering, advised that JPL had "some fairly sophisticated instrumentation and communication" expertise and facilities and the United States should aim for more than simply going back into orbit.[63] He advocated missions to the moon instead. [47]

Space Race Tepidity

However, the response indicated a lack of enthusiasm from key scientists and politicians about any endeavors to the moon stemming from a belief that such proposals were more stunt than science.[64] Even President Eisenhower looked down on the "useless" lunar probes and refused to be caught up in a "pathetic race" with the U.S.S.R.[65] [47]

Ironically, in November 1957, the Deputy Secretary of Defense Quarles publicly claimed there was "no cause for national alarm" from the success of Sputnik.[66] [47]

The Pioneer Program

The failure of Vanguard 1 on December 6, 1957, changed everything.

> *On Feb. 17, 1958, the Space Science Panel of Eisenhower's new President's Scientific Advisory Committee (PSAC)*[67] *— reorganized from the old Office of Defense Mobilization's Scientific Advisory Committee (ODMSAC) — held a meeting in the Executive Office Building that set two key objectives. Panel member Herbert York announced to attending representatives from JPL and Space Technology Laboratory (STL) that the*

[62] Support process observes and analyzes external events, then triggers event internal to system as option for decision makers

[63] Vision, concepts.

[64] Human nature.

[65] Human nature and politics.

[66] Misidentification of events, value of options/strategies.

[67] New decision maker.

> *committee had decided to attempt a lunar mission to make "contact of some type with the moon as soon as possible,"*[68] *with the stipulation that the contact had to have a significance "such that the public can admire it."*[69] *York said that the panel had concluded, given the second objective, that "some kind of visual reconnaissance,"*[70] *such as a camera to take a picture of the back of the moon, was the most significant experiment that a lunar vehicle could carry.* [47]

The following month, Eisenhower followed the PSAC endorsement and approved funding for five Pioneer lunar probes. On March 27, 1958, authorization for the one-year Pioneer program came from the new Advanced Research Projects Agency (ARPA). Of the five attempts, the first three were handed over to the Air Force to take advantage of the ready availability of its launch vehicles. The final two launches were under the direction of the Army, and therefore JPL.[71] [47]

Pioneer was publicly promoted by President Eisenhower[72] as a project "to determine our capability of exploring space in the vicinity of the moon, to obtain useful data concerning the moon, and provide a close look at the moon." [47]

> *The Pioneer program required simultaneous development of launch vehicles, spacecraft, and ground support stations. Crucial to the plan were the ground stations,*[73] *which would transmit commands to the spacecraft, determine their positions and instantaneous velocities, and receive data from them. Without them no close-up photograph of the moon could be received and, more fundamentally, no confirmation that the spacecraft were anywhere near the moon would be possible.* [47]

This system of ground stations, developed for the Pioneer program under the management of JPL and the visionary Eberhardt Rechtin, would eventually come to be known as the Deep Space Network.

---

[68] Objective 1
[69] Objective 2
[70] Requirement.
[71] Mission-level strategy with shaping factors; leverage legacy components, reuse.
[72] Public relations.
[73] Critical path.

2.1.4 Designing the Deep Space Network

The development of the future Deep Space Network was at a crossroads. Should the network design focus only on supporting the needs and limited objectives of the Pioneer program[74] or should the network be constructed to enable the likely missions of the future while meeting the immediate needs of Pioneer?[75] [47]

The Space Technology Laboratory Approach

The Air Force worked with the Space Technology Laboratory (STL) to prepare for the three Air Force Pioneer probes. The fast-paced timeline gave STL less than five months to set up a network, forcing the decision makers to focus exclusively on meeting the needs of the Pioneer program.[76] Station locations were chosen strictly for their favorable look angles[77] for transmitting commands to insert the probes into lunar orbit. An altered version of an antenna under construction[78] for the U.S. Air Force Discoverer reconnaissance satellites was installed at South Point, Hawaii, sporting a 60-foot diameter parabolic transmitting antenna.[47]

A bigger challenge for STL was to identify an antenna and a location for receiving data from the Pioneer probes. Photos of the moon would be sent back once the satellite achieved lunar orbit. This operational plan meant that a receiving antenna would need to be in the region of Europe and Africa as the spacecraft would be "passing over the prime meridian" during this critical time period. Furthermore, STL desired as large an antenna as possible to maximize the photo quality. Diplomatic, scheduling and funding issues constrained the team[79] to utilizing a pre-existing antenna in friendly territory. This antenna turned out to be a 250-foot (76-meter) diameter radio telescope that had been recently built by the University of Manchester at Jodrell Bank, England. Negotiations ended with STL being allowed to add a temporary feed and other equipment necessary to receive photos[80] from the Pioneer probes.[47]

STL continued using the 108-MHz operating frequency of the Vanguard and Explorer satellites.[81]

[74] Short-term strategy
[75] Long-term strategy: take a hit up-front to open future options
[76] Tight time constraint drives short-term strategy decision.
[77] Physical architecture decisions, location.
[78] Modified legacy design.
[79] Evaluation and decision constraints.
[80] Temporary technical modifications.
[81] Legacy.

Engineers at STL had the foresight to realize that the direction of space technology would drive the need for a permanent network of antennas. However, STL politics prevented the laboratory from taking an active role in the development of such a network.[82] By the time STL realized its mistake, JPL had already positioned itself to take the lead on the development of a deep space network.[83] [47]

The Vision of JPL

JPL's strategy was largely influenced by the brilliance of the visionary Eberhardt Rechtin,[84] who was chief of JPL's guidance research division. In 1958, while many scientists were pressuring for lunar missions, Rechtin argued for sending meteorological and surface-condition instruments to determine "the practicality of putting people on Mars", as he felt that Mars would be "one of the major goals of national prestige between the United States and the U.S.S.R." Scientists at JPL considered a planetary mission to be the ultimate engineering challenge. A permanent network of antennas was critical to this visionary program of exploration.[85] This Deep Space Network would be required to resolve spacecraft position and velocity as well as to send commands and to receive telemetry data.[47]

The Army/JPL Pioneer team had eight months to launch. The extra few months enabled them to build their own just-in-time network.[86] Thus, JPL took the long-term approach[87] to the antenna design:

> *The design of the stations should be on the basis of a long-term program. This means that the antennas should be precision built rather than simply crudely constructed telemetering antennas...it is much more practical in the long run to set up appropriate stations in the beginning of the space research program. The net cost will be much lower, flexibility of the program will be increased, and all program contractors can be served.* Eberhardt Rechtin, in a series of telex's to an Army Ballistics Missile Agency (ABMA) official in April 1958.[47]

[82] Human nature, politics.
[83] Competition; lost opportunity.
[84] The human factor: a visionary.
[85] Critical path.
[86] Opportunity.
[87] Strategy.

JPL's Network Design

Rechtin realized that his permanent network would have to serve two competing interests:[88] (1) continuously tracking the motion of space assets at (2) minimum cost. Geometry provides the answer. The optimal architecture occurs by separating three stations by 120° longitude.[47]

Next, Rechtin focused on designing the best possible communication system. He collaborated with the heads of JPL's electronics research section and the guidance techniques research section and determined that "it was important that the basic design be commensurate with the projected state of the art, specifically with respect to parametric and maser amplifiers, increased power and efficiency in space vehicle transmitters and future attitude-stabilized spacecraft." This strategy would allow the network to evolve into the envisioned permanent support system.[89] [47]

The antennas themselves had some rigid requirements:[90] a pointing accuracy of two minutes per arc or better to be maintained 24 hours a day, a structure robust to expansion and contraction of materials during sun exposure or ambient temperature variations, usable in winds up to 60 miles per hour, and able to endure winds up to 120 miles per hour while stowed. The antennas had the longest lead-time of any of the planned network's components. Rechtin demonstrated his incredible prescience by initiating the antenna design identification seven weeks before the Pioneer program was approved by Eisenhower. The task fell to William Merrick, head of JPL's antenna structures and optics group.[47]

JPL's plan was so ambitious that when Merrick consulted radio astronomers and suppliers, they "questioned our sanity, competence in the field and/or our ability to accomplish the scheduled date even on an around-the-clock effort." [91] [47]

Eliminating existing antenna designs[92] seemed to be modus operandi for Merrick. His reasons for rejecting designs included: foreign manufacture, cost, size, design flaws and construction time.[93] Tellingly, he automatically discarded the same Jodrell Banks antenna that STL choose for three important reasons: size, cost and time for development and construction (an incredible 7 years).[47]

The chosen design[94] was priced around $250,000 and met the requirements the team had compiled:

[88] Design trades.
[89] Planned for evolvability from the beginning.
[90] Design and performance requirements.
[91] Human nature.
[92] Decision method, strategy evaluation.
[93] Other considerations.
[94] Antenna design choice.

> *The 85-foot-diameter antenna had an equatorial mounting (one whose main rotational axis is parallel to the earth's axis) and this mounting was cantilevered for strength. Its unusually large drive gears for hour-angle (celestial longitude) and declination (celestial latitude) gave a high driving accuracy even though the teeth were not shaped with high precision; moreover, the sheer number of teeth meant that each tooth bore a low load even in high winds.* [47]

The antennas were available through Blaw Knox. The company had several other unrelated orders in the queue when JPL made their decision. The Army used its influence to move one of the three JPL antennas to the front of the line.[95] Having only one of three antennas manufactured on time was just as well. The planned overseas stations were hitting diplomatic hurdles and bureaucratic red tape,[96] and could not be completed by the second Army/JPL Pioneer probe. Fortunately, the requirements for the Pioneer program allowed JPL to make do with a single antenna placed in the United States. To compensate, JPL engineers designed the operations schedule so that the probes lunar arrival would coincide with the antenna's line of sight.[97] [47]

Furthermore, JPL's operations strategy mitigated a lot of the risk[98] associated with the STL program by not making an attempt to insert the probe into lunar orbit. Rather, JPL's probe would merely fly by the moon and would automatically take photos when the probe entered an appropriate range to the moon. This strategy also eliminated the need for an earth-based transmitter, thus buying the network team more time for building up their evolvable system.[99] [47]

The location of the United States station would be key to the future of the network. The further a spacecraft traveled from Earth, the weaker the received signal. Thus, this first site had some special requirements:[100] the antenna needed minimal outside radio interference, which could be accomplished by a natural bowl-shaped valley devoid of radio sources such as power lines, aircraft, and transmitters; stable soil to support the structure; an access road to transport materials; and it all had to be on government-owned land due to the imposed funding and time constraints.[101] JPL found its site near Goldstone Dry Lake in California. General Medaris had to use his influence to

[95] Influence.
[96] International politics.
[97] Scheduling.
[98] Risk mitigation.
[99] Mission drives architecture; architecture drives mission.
[100] Site requirements.
[101] Constraints.

secure Goldstone for the Pioneer program facilities, overruling another Army General who wanted the area at Camp Irwin in the Mojave Desert for use as a missile range.[102] [47]

A month before the first Army/JPL Pioneer probe in November 1958, the antenna at Goldstone passed its optical and radio-frequency tests and became operational. [47]

The team at JPL diverged further from the STL design by choosing a different operating frequency than the Vanguard and Explorer satellites.[103] Taking advantage of their opportunity to design the right system rather than constraining themselves to the legacy of Vanguard and Explorer, JPL engineers decided on an operating frequency of 960-Mega Hertz (MHz).[104] They based their decision largely on the fact that the growth potential of their network would be significantly limited below 500-MHz due to radio noise from terrestrial and galactic sources. [47]

Both STL's and JPL's systems, including several small antennas placed at the Cape Canaveral launch site as well as down range from it, performed adequately during the missions. Unfortunately, only the second Army/JPL probe, Pioneer 4, made it into space. To add insult to injury, Pioneer 4 missed the moon fly-by on March 4, 1959, passing too far for the camera system to activate.[105] Russia then launched Luna 3 on Oct. 4, 1959, successfully taking pictures of the far side of the moon.[106] [47]

JPL versus STL

Following Pioneer, JPL turned to expanding its ground support system into the envisioned global network. Part of this venture involved fending off a series of challenges from STL, the Deputy Secretary of Defense Quarles and the NRL. [47]

The first challenge came on June 27, 1958, when STL proposed a similar three-station network, involving their 250-foot-diameter antennas.[107] The Jodrell Banks-type antennas were to be built in Brazil, Hawaii and either Singapore or Ceylon. STL promoted a dual-network system, with stations spaced at 60° around the equator. STL's proposal did not indicate why two, three-station networks were necessary, simply stating "the estimates given here are believed to be realistic for completing construction of the first antenna in Hawaii in 16 months — by Oct. 15, 1959." The original Jodrell

[102] Politics and influence.
[103] Design divergence, "build new".
[104] Opportunity leveraged.
[105] Failures.
[106] International competition.
[107] STL threat; support process: option alternative.

Banks antenna took seven years to complete design and construction, so it was unclear how STL expected to meet the timeline. Furthermore, the estimated cost of the system was $34 million. Not surprisingly, the proposal went nowhere.[108] [47]

In early July, the separate ground support systems being developed by STL and JPL were challenged by Deputy Secretary of Defense Quarles.[109] Rechtin immediately headed to Washington, D.C., and convinced the chairman of an ARPA advisory committee on tracking, Richard Cesaro, that JPL's network deserved close attention. JPL was directed to submit a Proposal for Interplanetary Tracking Network.[110] The proposal had to meet the requirements of six ARPA reference programs. The July 25th proposal recommended a second tracking antenna at Woomera, Australia, and a third somewhere in Spain. Amazingly, the projected cost of JPL's network was under $6 million. Cesaro decided to recommend the Army/JPL to manage all of the space tracking and computational facilities.[111] [47]

JPL versus NRL

The battle for JPL's direction of the future Deep Space Network was not over. Rechtin anticipated a fight from the NRL which almost certainly thought it knew more about tracking than Army/JPL. Rechtin expressed his concern in an August 6 telex to a colleague, stating that Cesaro "may be overoptimistic" in believing ARPA would have sufficient influence to "put down any rebellion." Adding to his caution was the upcoming establishment of NASA on October 1. A civilian space agency meant that ARPA, as the interim space agency, would soon lose its political power. To complicate things further, the Department of Defense would soon desire its own tracking network due to secrecy concerns.[112] [47]

In late 1958, Rechtin's fears concerning the NRL came to fruition. The NRL Radio Tracking Branch was transferred to NASA, and as expected, its head John Mengel fought JPL's extensive plan for the support network.[113] Mengel argued that expanding the NRL's Minitrack network was more important to near-term American space interests than JPL's intended growth: "the satellite experiments and their associated tracking [were] more important than the deep space effort as far as NASA plans

[108] Unrealistic proposal for network architecture.
[109] Threat.
[110] Decision maker requests support process proposal.
[111] Influence.
[112] Human nature, politics, influence.
[113] Politics.

were concerned."[114] Fortunately for JPL, it had also been acquired by NASA by this point, and had built up some support.[115] NASA appreciated JPL's ideas for future lunar and planetary exploration, and had endorsed them since early November 1958. On July 10, 1959, NASA formally decided to move forward with JPL's plan.[47]

The Birth of the Deep Space Network

As NASA was a civilian agency, JPL could move toward South Africa as a host country.[116] South Africa was more optimal than Spain as most probes would pass over this region during the injection phase.[47]

Rechtin lobbied for local nationals as the operators for the overseas stations. He felt that international cooperation would encourage the best possible performance, particularly from professionals "proud of their work, held responsible, and cooperatively competitive in spirit" and "a bit of national pride certainly doesn't hurt!"[117] History would prove him correct.[47]

In collaboration with Australia's Weapons Research Establishment (WRE) and South Africa's National Institute for Telecommunications Research (NITR), JPL selected sites near Woomera, Australia and Johannesburg, South Africa. NASA endorsed the sites and construction began. Rechtin made sure that both WRE and NITR held responsibility for various key parts of the project to encourage their cooperation and continued participation.[118] [47]

The DSIF, consisting of the stations at Goldstone, Woomera, and Johannesburg, was operational in time to support the Ranger Program to acquire the first close-up images of the lunar surface beginning with the launch of Ranger 1 on August 23, 1961.[47]

In a memo sent out by JPL's Director, William Pickering, on December 24, 1963, the DSIF was formally redesignated as the Deep Space Network.

2.1.5 Analysis and Commentary

The history leading up to the Deep Space Network can teach us a lot about the interplay between the support process organizations and the decision makers. Examining the politics, influence and even

[114] Strategy: short-term gain vs. long-term advantage.
[115] Politics and influence.
[116] International politics.
[117] Human nature.
[118] Split responsibility and parallel process; human nature.

management styles of the Vanguard program and STL's later Minitrack proposals is very revealing. It is very clear in hindsight why JPL, Orbiter and the DSN were considerably more successful in the end than the NRL, Vanguard, and Minitrack.

Analysis of the DoD Proposals

A brief analysis of the DoD proposals highlights why it is no real surprise that the Air Force World Series was never a serious contention, Vanguard failed and Orbiter eventually succeeded.

The Air Force World Series proposal consisted of a prototype ICBM combined with existing components. The ICBM program was politically hot at the time. Not only was the ICBM program development crucial for continued success in the Cold War, but it was politically necessary that no IGY contributions use military rockets. Even if politics did not forbid the use of the ICBM, the fact that the rocket was a prototype introduced significant cost and technical risk. Overall, the Air Force proposal was dead before it ever really lived.

The Navy Vanguard proposal was a fairly complex, mostly new architecture. The design essentially created platforms of existing obsolete components. The existing designs were stretched to the required size and then fitted with appropriately sized engines. The third stage design required an entirely new engine to be designed — one that wasn't even prototyped yet! The Vanguard design appears to not only have required the most work and rework to implement and test, but it also likely incurred the most technical risk and the highest cost of all three proposals. Perhaps a clue of its future fate could be found in the use of obsolete designs?

The Army/JPL Orbiter proposal involved a simple architecture that relied heavily on leveraging existing infrastructure and components (i.e. strong reuse of tested and proven components). The architecture sported the least cost and technical risk and likely required the least work and rework to implement.

The obvious "best" decision based on the technical merits of the proposals was Orbiter. The choice of Vanguard over Orbiter was entirely a political decision. In this case, an over-emphasis on politics over technical merit resulted in the selection of the worst possible technical option. The eventual result therefore should not have been unexpected.

Analysis of Vanguard failure

Vanguard was set up for failure from the very beginning. It's politically-motivated selection overlooked its significant technical flaws. Arguably incompetent management of the development of Vanguard sealed its fate.

The design fundamentally assumed the availability of the Viking development team. Adding the Viking development team to the personnel roster therefore became a critical path item. Management of the program somehow missed the fact that the Viking team was off-limits since they had been reassigned to the new ICBM Titan program. This major oversight was likely the result of unwarranted assumptions and a failure to adequately research critical path risks.

Furthermore, the design itself appears to have been very sensitive to the payload size, which was known to be something that would be decided upon external to the chosen program. An unexpected payload size increase by the outside stakeholder forced a redesign of the second stage, adding even more development time and cost to the project.

Even more disturbing was the fact that the third rocket stage was not yet designed let alone built as a testable prototype. The management team attempted to mitigate the risks by contracting two companies in parallel, but in reality it was likely already too late.

There was no clear plan for the launch site. The assumed site ended up being untenable and the management team had to pull a few quick strings to get anything in place.

The project was so poorly managed that the budget quickly grew from $12 million to $100 million, requiring a financial save by the CIA.

Public expectations for the post-Sputnik test launches were inflated by the press, making the failure of Vanguard 1 a huge public embarrassment.

The Vanguard program had a very tight schedule and should have had a very tight budget to work within. The decisions made by the management team all but guaranteed that the schedule and financial budget would be stretched to the limit, and that the technical risks would be maximized. Fundamentally, the decision makers failed to clearly identify the critical path issues and the risks associated therein.

Analysis of the STL and JPL Approaches

Evaluating the differences between the approaches taken by the STL and JPL is very enlightening. The two organizations were in stiff competition during the Pioneer program, both attempting to position themselves for the design and construction of a permanent ground network. The STL worked with the Air Force for the first three Pioneer launches; JPL worked with the Army for the final two.

It is very clear from the history that politics and influence helped JPL tremendously while internal politics significantly hurt the STL. JPL benefited from the influence it had with General Medaris and other important government figures, who aided JPL in everything from site selection to manufacturing. Internal politics stemming from a desire for equity in the participation of various company divisions prevented STL from focusing their design on a permanent network until it was too late.

JPL benefited from having the right people in the right place at the right time. First and foremost was Rechtin, whose prescient vision and ability to leverage his keen understanding of human nature did the most to bring his evolvable deep space network to fruition.

Due to time constraints and the aforementioned internal politics, the STL went with the short-term approach by building its ground network solely focused on the requirements of Pioneer. History demonstrates that the team at STL was not very good at identifying the critical path issues and the associated risks. The most obvious example of this was the proposal to use Jodrell Banks type antennas for their permanent network. STL lost in the end. JPL on the other hand, went with the long-term strategic approach, positioning themselves early on to make the most of their resources, and won.

JPL had the advantage for several more reasons:

- The team was judicious with its choice of legacy over building new and vice versa.
- The team clearly identified threats and opportunities and immediately took steps to respond appropriately.
- Critical path items and their associated risks were clearly identified and dealt with.
- The team devised and implemented strategies to minimize cost and risk and to gain both the short- and long-term advantage.
- The team was responsive and adaptable to unexpected events.

In summary, it seems that even in hindsight, JPL did all of the right things at all of the right times at this point in the history of the deep space network.

2.2 The Evolution of the Deep Space Network

The evolution of the Deep Space Network can be broken down into four levels.

- Change within and between the organizations comprising the DSN.
- The increasing number and complexity of missions.
- Changes in the composition of the physical architecture of the DSN.
- Improvements in the underlying technology of the DSN.

This section details and analyzes the evolution of the DSN within and between each of these four levels.

2.2.1 Organizational Evolution

The organizational evolution of the Deep Space Network proceeded in three distinct stages as shown in Figure 2.1. This section highlights the key organizational changes and overall trends within and between each of these stages.

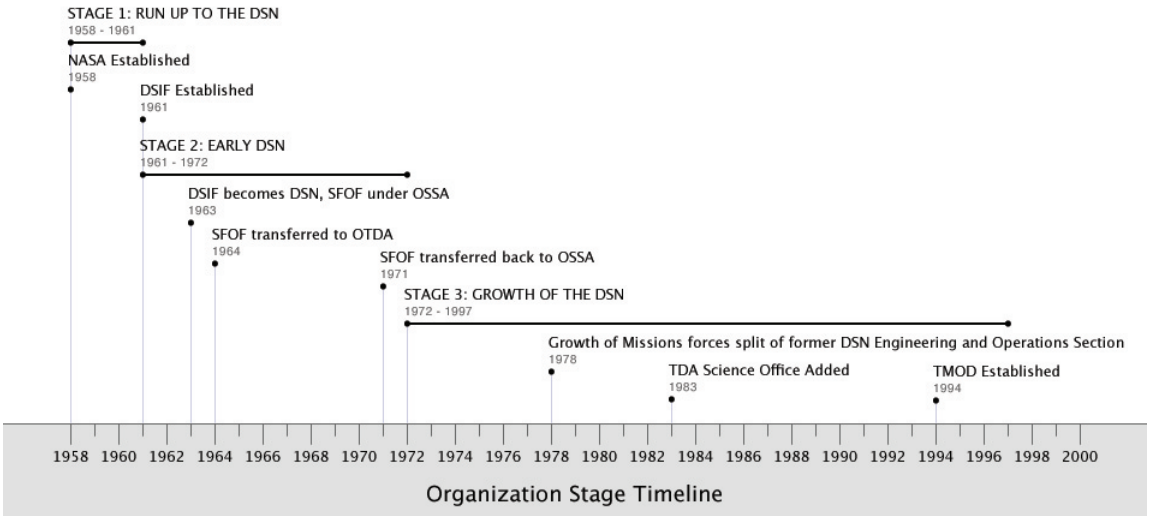


Figure 2.1: DSN organizational evolution timeline. The DSN proceeded in three distinct stages.

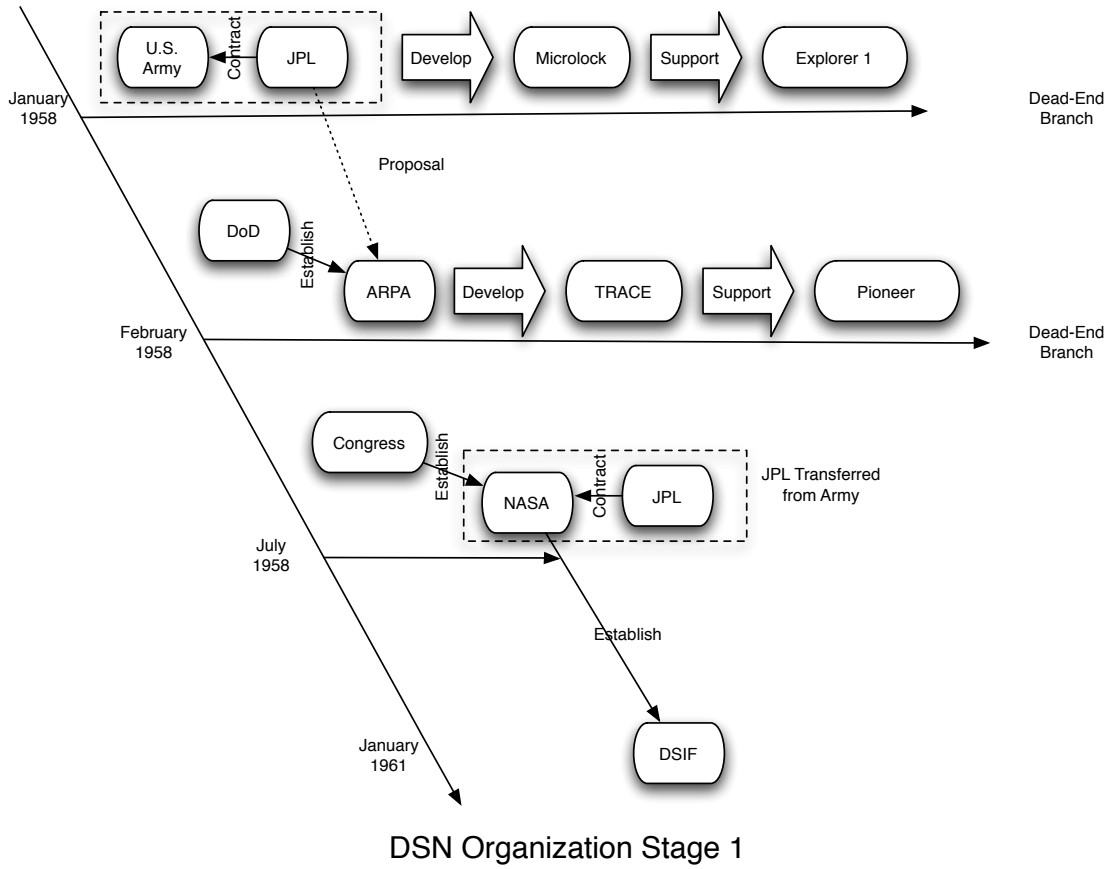


Figure 2.2: DSN organization Stage 1.

DSN Organization Stage 1

The first organization stage occurred very early on, starting five years before the birth of the DSN. Several organizations and ground network combinations were tried before the United States settled on the Deep Space Instrumentation Facility DSIF under NASA/JPL supervision. Figure 2.2 provides a depiction of this stage of the organizational evolution.

In January 1958, the U.S. Army with JPL as an independent contractor worked on developing the Microlock network for the Explorer 1 mission. It was clear, however, that the network would be insufficient to support the Pioneer program requiring tracking at lunar distances.

In February, the (DoD) established the Advanced Research Projects Agency ARPA. ARPA was assigned to oversee the Pioneer program. In this capacity, the organization approved a JPL plan for a network of 26-meter tracking antennas that ARPA planned to develop as the Tracking and Communications Extraterrestrial Network (TRACE) network. TRACE would thus be used to support

Pioneer.[48]

In July 1958, Congress established the National Aeronautics and Space Administration (NASA). The civilian space program as well as JPL were soon transferred over to NASA. At the time, the first TRACE antenna was under construction. Under NASA, this antenna was renamed Pioneer Station.[48]

When the DSIF was formed in January 1961, Pioneer Station was designated DSIF-11.[1]

DSN Organization Stage 2

The second stage follows the development of the early DSN. The organization largely remained the same from 1963 to 1972 with the exception of settling on who was responsible for the SFOF. The 1972 Viking support system required a temporary organizational change. See Figure 2.3 for a graphical representation of the changes described below.

Eberhardt Rechtin was named the Director of DSIF when it was formed in January 1961. Funding and oversight was jointly maintained by JPL's TDA office and NASA's OTDA.[1]

In December of 1963, Pickering established the DSN by combining the existing DSIF (now known as the TDA Program Office), the Intersite communications grid, and the mission-independent portion of the SFOF at JPL. The SFOF was under construction at the time, but was funded by the NASA Office of Space Science and Applications OSSA via JPL's Lunar and Planetary Projects Office LPPO. The SFOF was completed in October.[1]

The following year, the Intersite communications became known as the Ground Communications Facility GCF and responsibility for the SFOF transferred from OSSA to OTDA. The next years saw a rapid increase in the number of complexity of missions.[1]

Finally, in 1971, the OSSA was transferred back to OSSA from OTDA:

> *The deep space missions of the late 1960s and early 1970s brought with them substantial increases in the SFOF data-processing load. Much of this data processing was strictly scientific and unrelated to DSN operations. Yet, the tracking and data acquisition function was obligated to provide for this computer time without the authority to review requirements. The flight projects were, in essence, requesting and getting large blocks of computer time and were neither financially nor managerially accountable for*

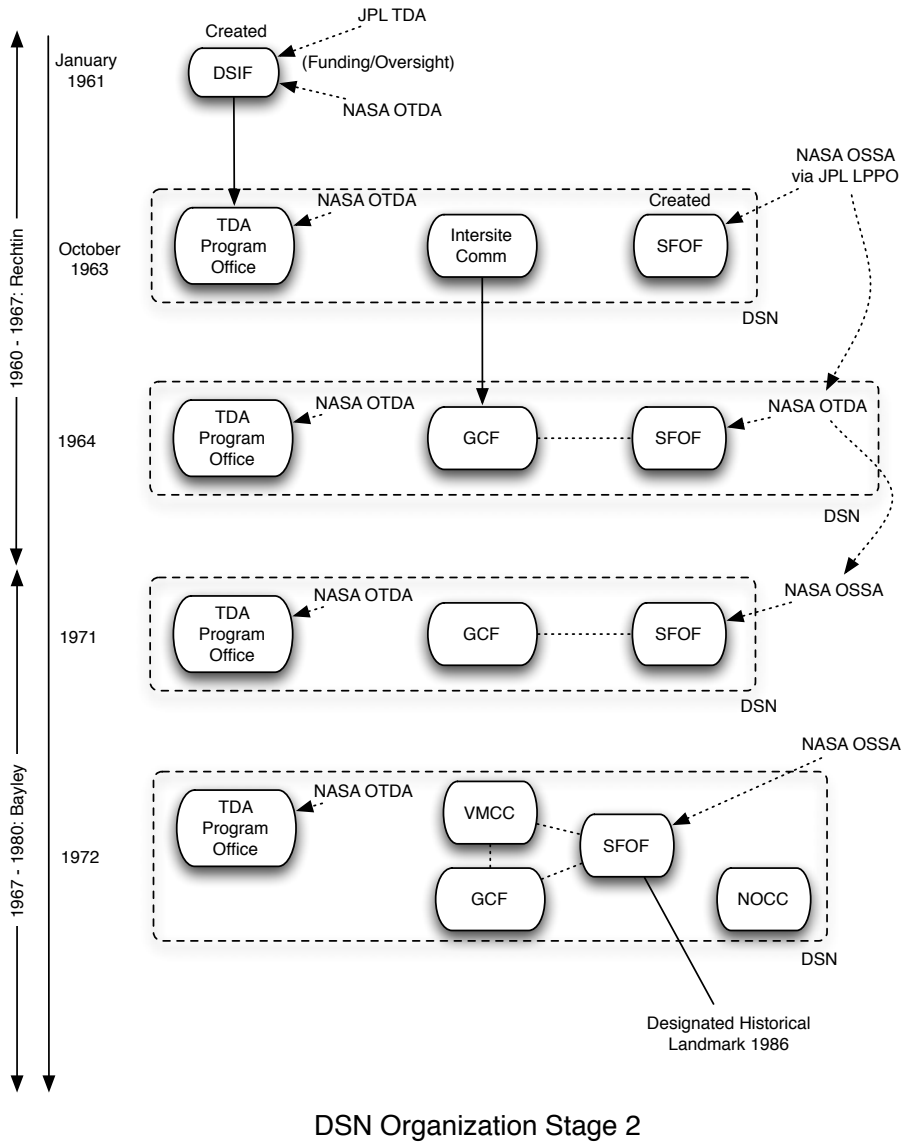


Figure 2.3: DSN organization Stage 2. For reference: Deep Space Network (DSN), Deep Space Instrumentation Facility (DSIF), Tracking and Data Acquisitions (TDA), Office of Tracking and Data Acquisitions (OTDA), Space Flight Operations Facility (SFOF), Office of Space Science and Applications (OSSA), Lunar and Planetary Projects Office (LPPO), Ground Communications Facility (GCF), Viking Mission Control Center (VMCC), Network Operations Control Center (NOCC).

them. It was a bad managerial situation. NASA Headquarters recognized the situation and, in October 1971, Gerald Truszynski OTDA and John Naugle OSSA reviewed the problem and decided to transfer the SFOF functions from OTDA back to OSSA. In this way the responsibility for review and validation of requirements and the associated

costs of scientific data processing would be borne by the flight projects themselves. [1]

Although the separation of the SFOF from the DSN may have appeared to be merely a "paper exercise," it was not accomplished without considerable disruption to the carefully crafted interface agreements already in place between the DSN and the Pioneer and Viking flight projects. Schedules, interface agreements, and capabilities had been negotiated with various elements of the flight projects and had been formally documented and approved, in accordance with current practices. These schedules, agreements, and capabilities, of course, included the SFOF as well as the DSIF. When the separation took place, new interfaces between the DSN and the flight project, and the DSN and Office of Computing and Information Systems (OCIS), had to be developed and documented. [1]

The Viking missions in 1972 saw a temporary change in the organizational structure:

The Viking Mission Control Center VMCC, which included the SFOF central computing system, the mission support areas, and the Viking mission simulation system, was the joint responsibility of the OCIS and the Viking Mission Operations System. The DSN was responsible for the deep space stations, which included the 64-meter and 26-meter subnets, and transport of data to and from the VMCC via the high-speed and wide-band data lines of the GCF. Control and monitoring of network performance and validation of the data streams flowing between the VMCC and the deep space stations was to be accomplished by a separate data-processing capability that would be independent of the mission-related computers in the SFOF. These functions would be accommodated in a new Network Operations Control Center NOCC, which was being designed at the time (1972). [1] p. 600

DSN Organization Stage 3

The third stage demonstrated substantial evolution within the Tracking and Data Acquisitions portion of the DSN due to the rapid increase in the number and complexity of missions. (See Figures 2.4 and 2.5 for five snapshots of the organizational charts for the TDA and Telecommunications and Mission Operations Directorate (TMOD), as the TDA became known in 1994.)

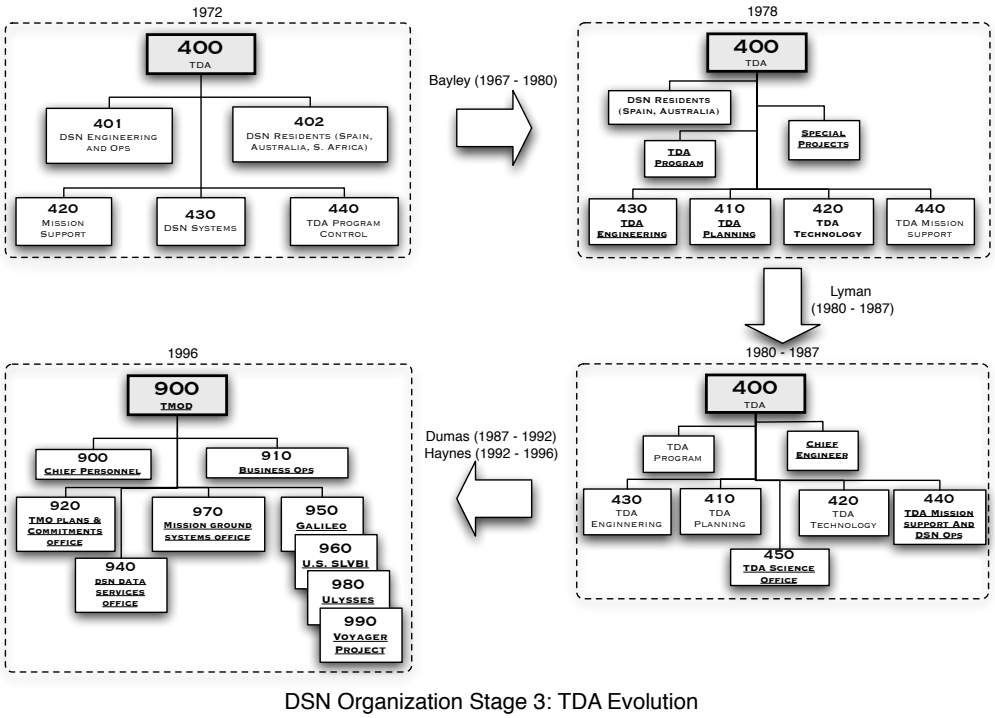


Figure 2.4: DSN organization Stage 3: TDA evolution. Figure adapted from Uplink-Downlink.

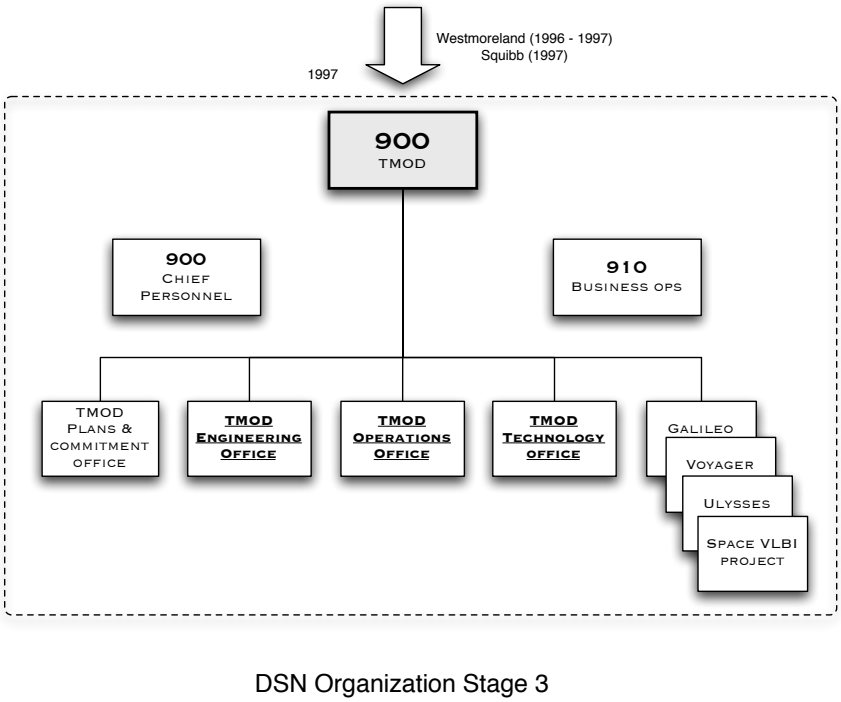


Figure 2.5: DSN organization Stage 3: TMOD. Figured adapted from Uplink-Downlink.

The first organization chart shows the TDA office as it was at the beginning of Stage 3 in 1972. By 1978, the DSN TDA office had out-grown itself, expanding to an incredible 74 people. To accommodate the boom of DSN flight operations, the former Engineering and Operations Section was split into the DSN Mission Support Office and the DSN TDA Engineering Office. The new Engineering Office, led by Renzetti, was charged with developing the DSN engineering systems. Spaulding took charge of the new Mission Support Office, responsible for all aspects of flight mission support excluding maintenance and active operations. These responsibilities included configuration control, "interfacing with the flight projects and scheduling antenna time."[1]

The TDA organization continued to grow considerably during the Lyman years (1980-1987). The TDA Science Office was added in 1983, including "a Geodynamics program, the Search for Extraterrestrial Intelligence (SETI) program, the Goldstone Solar System Radar program and several other special research projects." In 1986, the SFOF was designated a Historical Landmark by the U.S. Department of the Interior. The responsibilities of the TDA Engineering Office were expanded to include "interagency arraying, compatibility and contingency planning, and implementation of new engineering capability into the network and GCF."[1]

The organizational structure of JPL underwent significant changes during the Haynes years (1992-1996). The JPL organization in 1992 was headed by the JPL director who oversaw several offices run by Assistant Laboratory Directors (ALDs). At the close of Haynes' tenure, JPL was composed of many directorates led by separate directors. In the spirit of the new "policy for change," each of the directorates were also internally reorganized.[1]

More significantly, the Telecommunications and Mission Operations Directorate was established in 1994 to support the NASA Space Communications and Operations Program, which was part of the new leaner, cost-effective program instituted by then President Bill Clinton.[1]

The TMOD restructuring is described in Uplink-Downlink as follows:

> *Essentially, the former TDA organization was condensed into two offices: one for planning, committing, and allocating DSN resources; the other for DSN operations and system engineering. DSN science and technology were incorporated in the former, DSN development in the latter. In addition to these two offices, the Multimission Ground Systems Office, the project offices of the four inflight missions (Galileo, Space Very Long Baseline Interferometry (VLBI), Ulysses, and Voyager) and a new business of-*

> *fice were added to create TMOD. There could be little doubt that the TMOD was now operations-driven rather than engineering-driven.*

By March 1995, the Reengineering Team had completed its redesign of key subprocesses within the TMOD. In 1997, the TMOD was fully transitioned to the new process-based management structure. The allocation of resources and the new Customer Services Fulfillment Process would be managed out of the TMOD Operations Office, which was comprised of the previous DSN Data Services and Multimission Ground Systems Offices. A new TMOD Engineering Office was created for developing the "new system engineering functions" for the fulfillment process, including the asset creation process. The TMOD Technology Office was responsible for providing enabling technology. The remaining TMOD offices were largely left untouched.

Before TMOD, each flight project was assigned a TDA office representative to negotiate the use of the necessary tracking and data acquisition services. When the TDA office evolved into TMOD, the role of the DSN manager also changed. TMOD became "process-oriented," so it was a natural extension to expand the scope of the Tracking and Data System (TDS) representative beyond the interface of the DSN and the Multimission Ground Data System (MGDS) to include the whole Customer Fulfillment Process. In effect, the TDS manager would become a version of the "empowered customer service representatives."

2.2.2 Mission Evolution

To understand the evolution of the missions undertaken by the DSN, it's enlightening to consider their complexity in terms of distance from Earth and the mission "stage."

A table of the mission evolution as a function of the two kinds of complexity is depicted in Figure 2.6 on page 100. For each combination of mission descriptors (mission stage/distance from earth), the year is provided for the first time that kind of mission was successful. The temporal decades are color-coded. There are several assumptions made when considering the mission complexity. It is assumed that complexity increases the further missions occur from Earth, manned missions are more complex than unmanned and the mission stages are of increasing complexity.

It is clear from the mission complexity table that missions have been focused towards achieving early stages at all distances from Earth. This result implies that there are greater challenges to the spacecraft design and the DSN configuration with increasing complexity in mission stages (i.e.,

MISSION COMPLEXITY TABLE			Mission Stages										
			Probe Flyby	Probe Orbit	Probe Impact	Probe Land	ProbeLiftoff	Probe Surface Explore	Probe Base	Probe Base w/ Samples	Manned Orbit	Manned Land, Explore, Liftoff	Manned Base
Location with Distance from Earth			P1A	P1B	P2	P3A	P3B	P3C	P4A	P4B	M1	M3	M4
Inner System	Earth	0 AU	x	1958							1962		
	Moon	~0 AU	1959	1966	1964	1966	1967				1968	1969	
	Sun	1 AU	x	1960									
	Venus	0.28 AU	1962	1990	1994								
	Mars	0.5 AU	1965	1971		1976		1997					
	Mercury	0.61 AU	1974	(2011)									
Asteroid Belt			1973										
Outer System	Jupiter	4.2 AU	1973	1995	1995								
	Saturn	8.5 AU	1979	2004		2005							
	Uranus	18.2 AU	1986										
	Neptune	29.1 AU	1989										
	Pluto	38.5 AU	(2015)										
Transient Small Body	Comet		1985										
	Asteroid		1991	2000		2001							
Interstellar			2003										

Figure 2.6: DSN mission evolution as a function of complexity. The decades are identified by color; the darker the color the later the decade. Future missions are purple and are further distinguished with parentheses.

probe to manned missions) than for traveling further from Earth. Also, the further one travels away from the Earth, the more difficult it is to advance along the mission stages. Only the highest value locations (e.g., moon and Mars) have been targeted for increases in mission stage complexity so far.

Figure 2.7 presents a flowchart of the stages as derived from information on the missions attempted over the DSN lifetime. There are two fundamental types of missions: manned and unmanned. The four stages are: flyby/orbit, impact, land/explore/liftoff, and base. Based on actual missions, certain unmanned probe missions must be undertaken prior to the manned versions. This order is due to safety concerns for the astronauts.

Figure 2.8 on page 102 shows the evolution of mission complexity over time, broken into eras of mission complexity type (e.g., Probe Stage 1 for the Inner Solar System).

There is a clear progression in the mission complexity for the DSN. Considering only the unmanned probes, the inner system missions precede the outer system trips, and within each of these Stage 1 is followed by Stage 2, which is then followed by Stage 3. The inner system manned missions occur in a time period that spans portions of all three stages of the probe missions, corresponding to the fact that key operations and technologies were tested with probes before attempting similar missions with astronauts.

The outer system Stage 1 era has gone on for a very long time owing to the extremely long distances

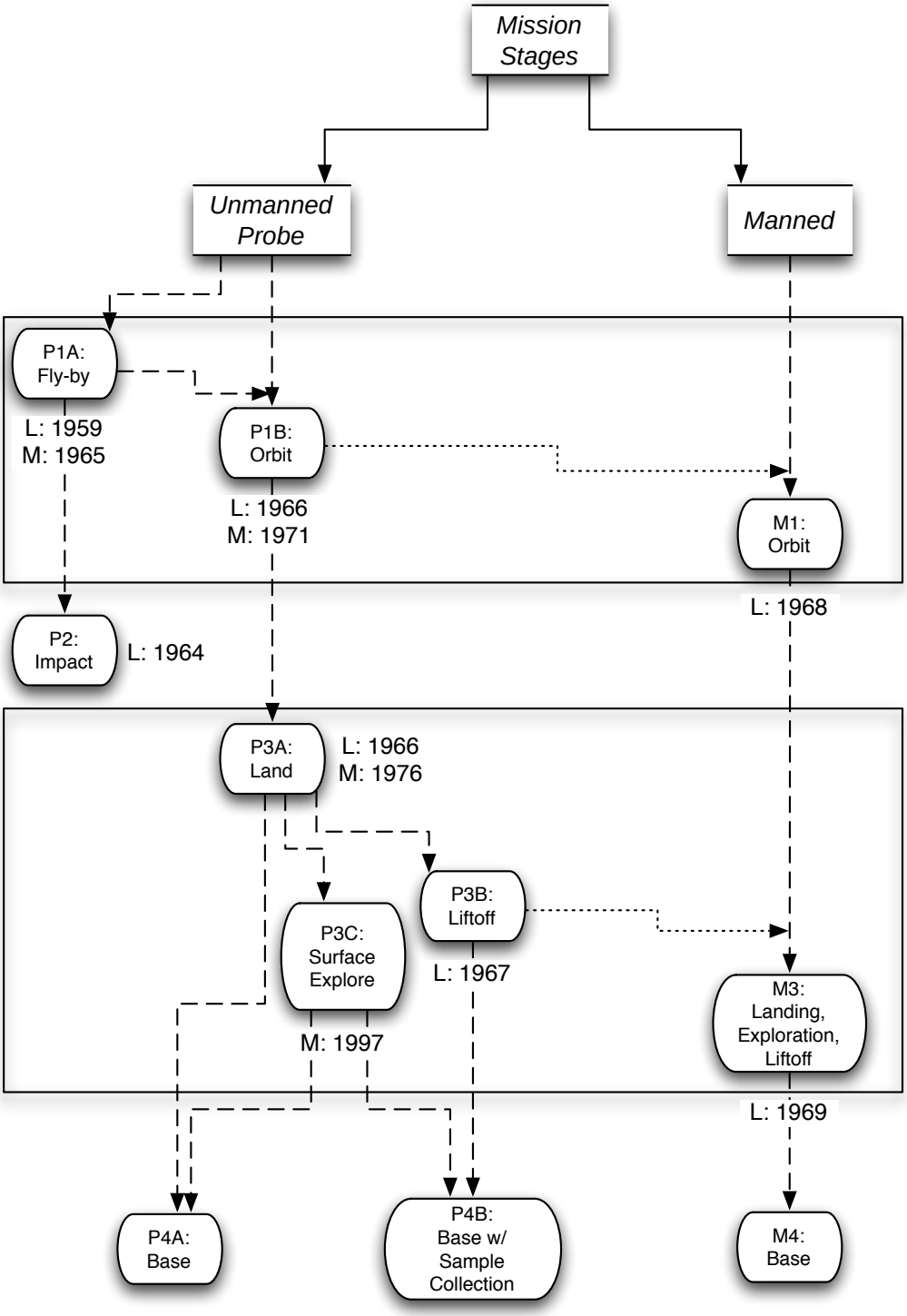


Figure 2.7: DSN mission stages for unmanned probes and manned missions. Example actual lunar missions are designated by 'L' and the year in which it occurred. Similarly, actual Mars missions are designated by 'M'.

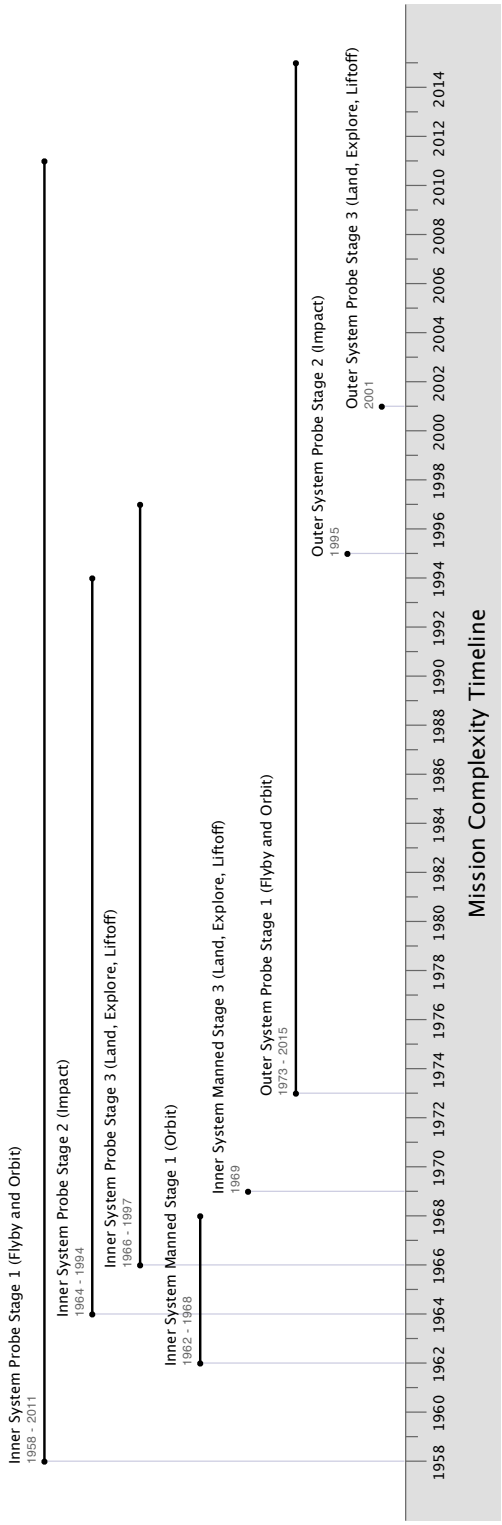


Figure 2.8: DSN mission complexity timeline. The figure shows the evolution of mission complexity over time, broken into eras of mission complexity type (e.g., Probe Stage 1 for the inner solar system).

the spacecraft must travel. The inner system Stage 1 has gone on for even longer, but the reasons are more subtle. The mission complexity table and the time line fail to show the multiple "rounds" of Stage 1 missions that have occurred. As technology has progressed and scientific interests wandered, different types of missions were sent out around the inner system. Some missions looked for signs of pre-existing or current life, some missions explored whether resources existed to support human bases, while others went to take advantage of the advent of mapping technology.

The data for Figures 2.6 on page 100 and 2.7 on page 101 are provided in B.2 as tables, which describe all of the important DSN missions with mission highlights, arranged by date. Also provided are the corresponding "first successful" mission stages as correlated in the Mission Complexity Table.

2.2.3 Physical Architecture Evolution

The evolution of the physical DSN architecture covers changes to the station complexes and the stations themselves (i.e., the network assets). This breakdown is reflected in the change taxonomy for the DSN physical architecture, as shown in Figure 2.9.

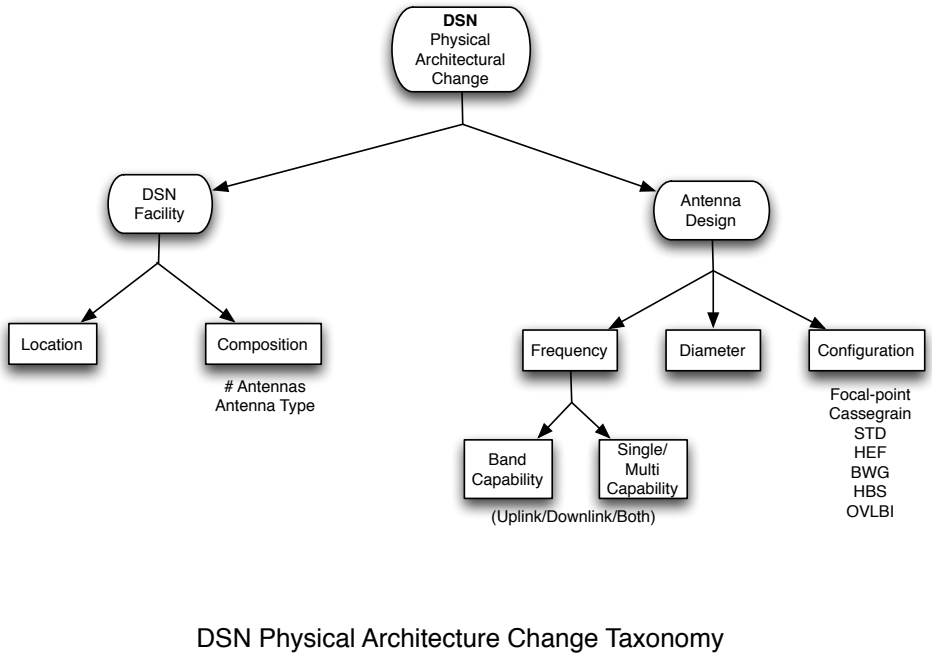
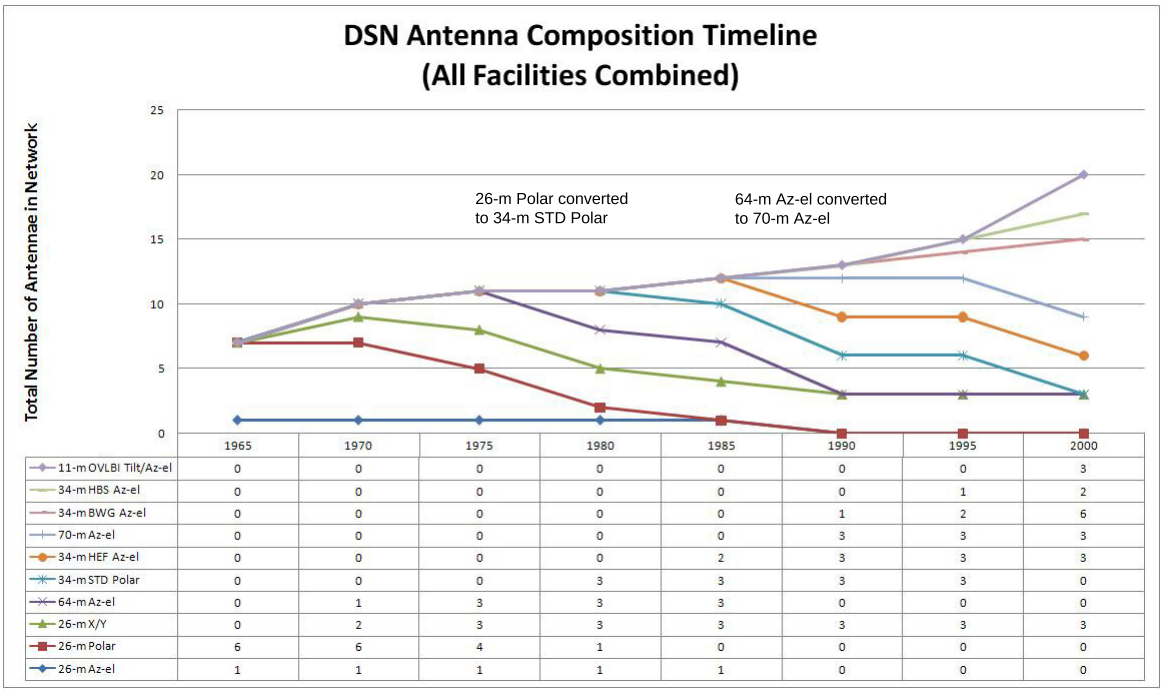


Figure 2.9: DSN physical architecture change taxonomy. The evolution of the physical DSN architecture covers changes to the station complexes and the stations themselves (i.e., the network assets).



	1965	1970	1975	1980	1985	1990	1995	2000
11-m OVLBI Tilt/Az-el	0	0	0	0	0	0	0	3
34-m HBS Az-el	0	0	0	0	0	0	1	2
34-m BWG Az-el	0	0	0	0	0	1	2	6
70-m Az-el	0	0	0	0	0	3	3	3
34-m HEF Az-el	0	0	0	0	2	3	3	3
34-m STD Polar	0	0	0	3	3	3	3	0
64-m Az-el	0	1	3	3	3	0	0	0
26-m X/Y	0	2	3	3	3	3	3	3
26-m Polar	6	6	4	1	0	0	0	0
26-m Az-el	1	1	1	1	1	0	0	0

Figure 2.10: DSN antenna configuration evolution for all complexes combined. STD: Standard; HSB: High-speed Beam Waveguide; HEF: High-efficiency; BWG: Beam Waveguide.

The station complexes can change location and composition over time. Location changes were rare, and only occurred in the early years of the DSN. The original DSN network was composed of complexes at Goldstone, California, Woomera, Australia and Johannesburg, S. Africa. As the number and complexity of missions expanded, the need for tracking multiple antennas grew. It was decided to build a second network consisting of overseas stations at Canberra, Australia, and Madrid, Spain. The initial overseas complexes were closed during a period of network consolidation in the early 1970's and operations were fully ceded to the Canberra and Madrid complexes.

Composition changes were much more common in the DSN. The number of each type of antenna changed every few years as more antennas were acquired, some were retired and others were converted to new uses. These changes are captured for the DSN network as a whole in Figure 2.10. The antennas are distinguished by their diameters (in meters), the type of mounting (Azimuth-Elevation (Az-el), Polar, X/Y, and Tilt/Az-el), and their configuration (e.g., STD, HEF, BWG, HSB, and Orbiting Very Large Baseline Interferometer (OVLBI)) when applicable.

The chart depicts several key things about the evolution of the physical architecture. New types of antennas have been acquired, legacy antennas have been converted to higher-performance antennas, antennas have been retired, and subnets have been expanded in number. For example, although

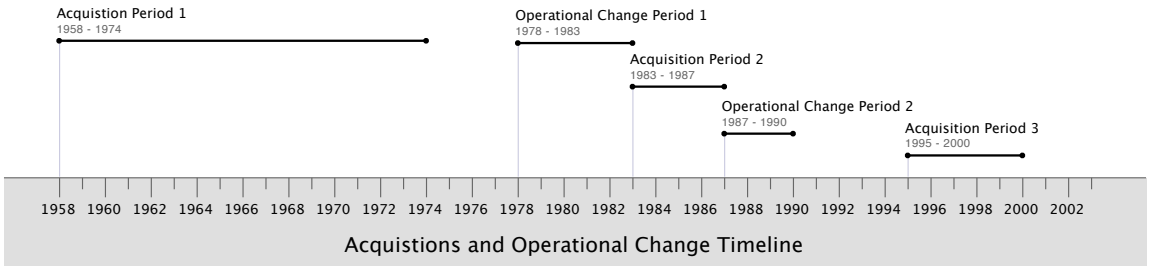


Figure 2.11: DSN acquisition and operational change timeline. Periods of network acquisitions of antenna stations historically have alternated with periods of changes made to legacy assets during their operation (e.g., expand 64-meter antennas to 70-meter). These cycles roughly correspond to intervals of economic downturns or crises.

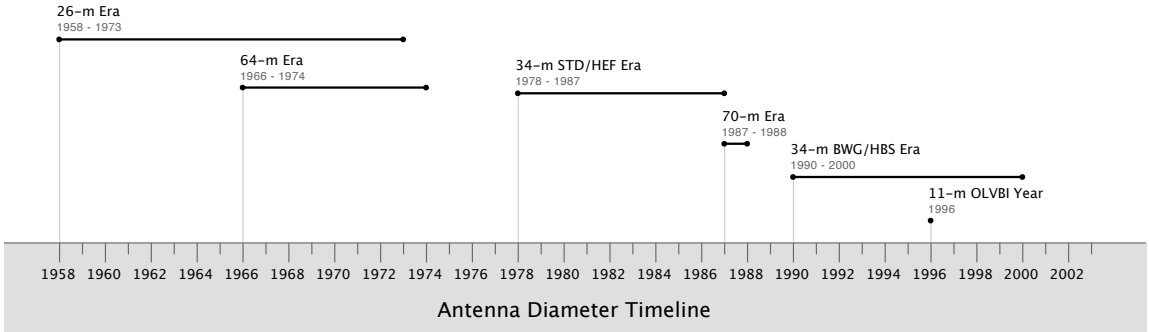


Figure 2.12: DSN antenna diameter evolution timeline.

the 26-meter antennas have been a part of the network the entire life of the DSN, a subnet of the Polar-mounted antennas were expanded to 34-meter by 1980. The first 64-meter antenna was acquired starting around 1970, with the full subnet coming into service by 1975. The 64-meter subnet was then rehabilitated and upgraded to 70-meter by 1988. The 34-meter STD Polar subnet was subsequently retired in 1999.

Periods of network acquisitions of antenna stations historically have alternated with periods of changes made to legacy assets during their operation (e.g., expand 64-meter antennas to 70-meter). See Figure 2.11 for a time line. These cycles roughly correspond to intervals of economic downturns or crises.

Figure 2.12 similarly depicts a timeline of the evolution of the station antenna diameter. During the first acquisition period, the DSN built or otherwise acquired many 26-meter antennas. Less than 10 years after the establishment of the DSIF, the DSN added a single subnet of 64-meter antennas,

ACQUISITION	OPERATIONAL CHANGE	OBSOLESCENCE
Build using modified COTS design	Add capability	Transfer out (network consolidation)
Build using scaled version of original COTS design	Increase diameter	Sell for scrap
Transfer in (network consolidation)	Repair/rehabilitation	Designate a historical landmark
Build using new design (with legacy commonality constraint)	Retrofit	Decommission
Relocate (network consolidation)	Replace	

Table 2.1: Change mechanisms for the DSN physical architecture evolution.

which were a scaled version of the 26-meter Polar antennas. During the subsequent first operational change period, the DSN extended a subnet of the 26-meter antennas to 34-meter STD. Acquisition period 2 saw the addition of a 34-meter HEF subnet. The 64-meter subnet was extended to 70-meter during Operational Change 2. An experimental 34-meter BWG was installed at Goldstone at end of that period. Finally, Acquisition 3 was a period of considerable growth, when the DSN acquired many 34-meter BWG and HSB antennas as well as a 11-meter OVLBI subnet.

Table 2.1 lists the different mechanisms of change for the evolution of the physical DSN architecture. These mechanisms are broken into three categories: acquisition of assets, changes to the assets during the operational phase, and changes resulting in the obsolescence of the assets.

Appendix B.3 contains detailed timelines that were used for the analyses presented here.

There are a few important things to note to fully understand the physical architecture evolution of the DSN. First, the early generations of antennas were based on the COTS design of the initial DSIF-11 (Pioneer Station) antenna. The first generation antennas were either identical, had modifications to the mounts, or were a scaled version of the DSIF-11 antenna. Second, later antenna generations can be traced back to Pioneer Station as the new designs were constrained to ensure commonality of components for maintenance, repair and training purposes.

Thus, the initial design decisions surrounding the Pioneer Station have affected every build decision since. This architecture legacy demonstrates change-resistance in terms of parts, training, knowledge-base and experience. Small deviations from the original design seem to be acceptable, but there were no instances of any "radical" change. This historical realization should serve to underscore the importance of legacy in complex systems.

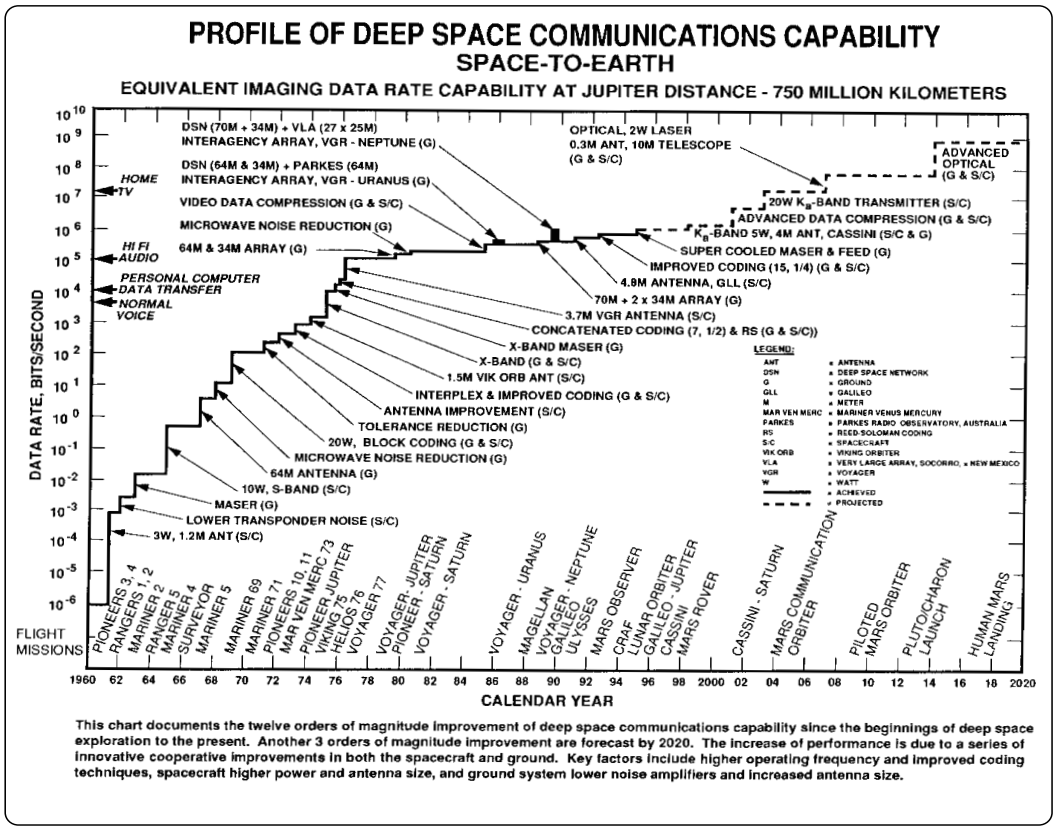


Figure 2.13: DSN technology evolution. Figure taken from Uplink-Downlink.[1]

2.2.4 Technological Evolution

The Technology evolution of the DSN is the most fundamental level of change. Technology feeds into every one of the higher levels and is similarly driven by them. The majority of technological changes take place at the component level, but several are at the physical asset and operational level (e.g., arraying antenna subnets to temporarily boost performance).

Changes in technology are the easiest type of change to correlate with measurable performance improvement, as Figure 2.13 demonstrates. The Profile of Deep Space Communications Capability chart[1] provides a graphical depiction of the evolution of technological advances and their corresponding improvements in equivalent imaging data rate capability at normalized Jupiter distance. Many of these changes were driven by increasing requirements stemming from missions of increasing complexity, while some technological advances enabled more complex missions. The performance of the technical changes flattens out over time (it is important to note that the y-axis is

IMPORTANCE	CHANGE	NOTES
Very important	1. SC0	1. Antenna Power. Important early on. Increasingly marginal returns as design evolved.
Important	1. G1 2. GSC1	1. Antenna Size. One data point. 2. Frequency. Backed by Anecdotal evidence. Chart missing some data.
Mild	1. SC1 2. G3 3. G4 4. G5 5. G6 6. GSC0	1. Antenna Size. Early design impact confounded with power effect. Seemingly impacted by other system changes. 2. Noise Reduction. About 60% improvement over 2 data points. Tolerance Reduction. 4. Microwave Amplification By Stimulated Emission of Radiation (MASER) (amplifiers). Decreasing impact as design evolved. 5. Arrays. Seemingly impacted by other system changes. 6. Coding, Compression. Effect trending downward, relative impact varies.
Low	1. SC2 2. SC3	1. Antenna Improvement. Only one data point. 2. Noise Reduction. Only one data point.

Table 2.2: Importance factors of types of technology change in the DSN. The table provides a breakdown of the types of technology change and the apparent relative importance to the communications capability improvement.

on a log scale). It seems to become more and more difficult to achieve a large boost in performance relative to previous changes. A portion of this trend may be an artifact of legacy. There may only be so much that advances in new technology can give for a fixed underlying architecture (the DSN has been and is currently a point-to-point ground-centric network).

Figure 2.14 shows a taxonomy of the DSN technological evolution. The changes in Figure 2.13 on page 107 are separated into three main categories based on where the change is made: spacecraft, ground and spacecraft, and ground. Identifiable subcategories of technological change are designated by a code. There are few data points, but these suggest several apparent trends.

Table 2.2 provides a breakdown of the types of technology change and the apparent relative importance to the communications capability improvement. The notes column highlights some of the trends. Overall, it is apparent that the impact of each change within each subcategory suffers from increasingly marginal returns in performance. There are several subcategories of change that seem to be impacted more significantly by recent changes in other areas. For example, the shift to X-band frequency in 1975 may have positively impacted the performance improvement when the spacecraft antenna size was increased a few years later.

By comparing ground and spacecraft changes in Table 2.2, it would seem that similar changes to the ground are on the whole more important than those on the spacecraft. There is not much data for GSC changes, but they seem more important than some adjustments to the spacecraft alone.

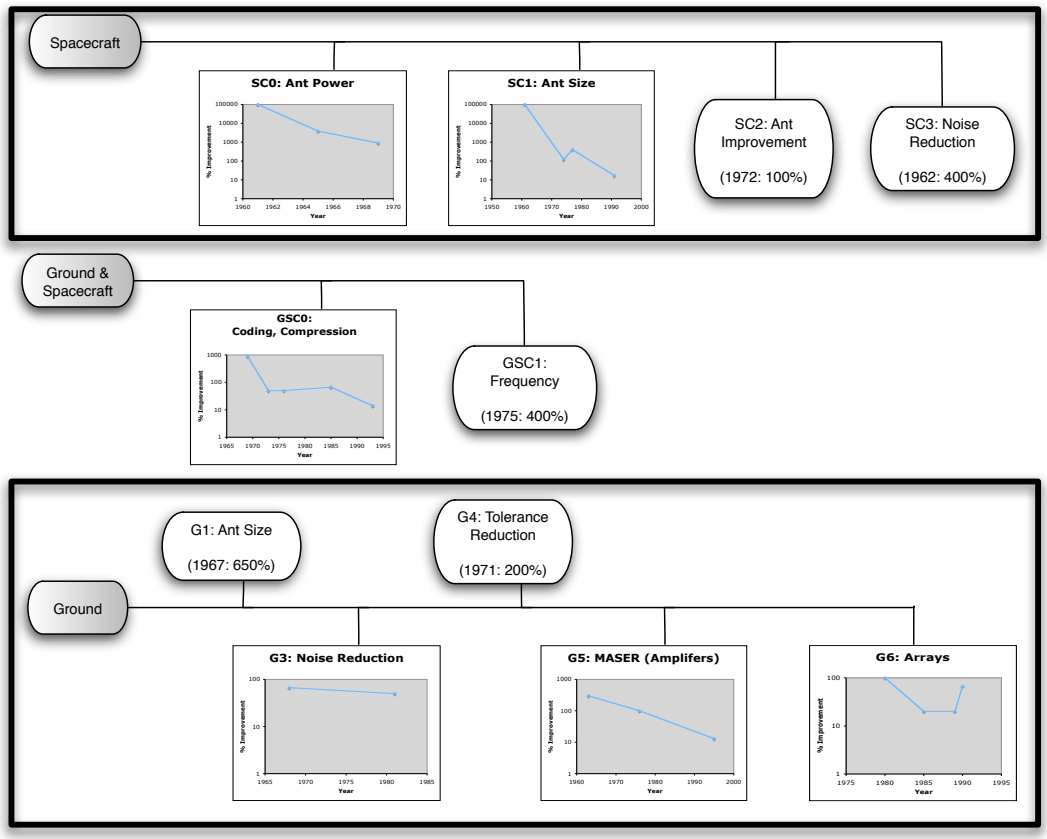


Figure 2.14: DSN technological evolution taxonomy. Technological changes are separated into three main categories based on where the change is made: spacecraft, ground and spacecraft, and ground.

Changes to assets on the ground seem to be very important. If this holds true, then building the right ground network is absolutely critical to the achievable performance of the system as a a whole.

2.2.5 Analysis and Commentary

Let's quickly review the main points that have been gleaned from collating information on the various levels of evolution for the Deep Space Network.

Organization

In the Organizational Stage 1, the decision makers had a short time frame to choose an initial network and establish the funding and oversight structure. There were dead-ends as a few ideas were

tried and rejected. The decision of establishing a new structure over continuing with an existing one was largely political and influence driven.

Stage 2 occurred over a medium time frame, which roughly corresponded to the Space Race with the U.S.S.R. (1957-1975). Organizations were created and consolidated to form the original structure of the DSN, which largely remained the same with the exception of sorting out funding and oversight authority for the SFOF and some internal reorganization to support Viking. It is likely that the decision makers minimized the organizational change during this period to focus resources on the Space Race.

In Stage 3, there was a rapid expansion of programs, with an increasing complexity and number of missions. This tremendous growth required more personnel involved in more responsive support processes. The personnel was largely comprised of a new generation of people with different attitudes from the founding generation. This era also benefited from a relaxation in the political climate that allowed more resources to go toward restructuring and moving the organization toward customer-centric processes.

Mission

At the mission level, there was a distinct focus on missions whose complexity was due to distance from the Earth. These missions were most likely chosen as they would provide the most rapid advancement of knowledge. The strategy appears to be one of scientific breadth: learn a little about each planet, the asteroids and comets, etc., then return later for more detailed information as time, money and technology permits. Only the highest value sites (e.g., the Moon and Mars) were targeted for tackling the advanced mission stages. This is a strategy of depth. The choice of highest value sites appear due to public interest, the most likely to use for future staging and most likely to enable successful bases.

Physical architecture

The cycles of acquisition and asset modification of the physical architecture correspond approximately with periods of economic downturn or crises. The system was able to achieve gains in performance or operations by adding assets during good economic times. During bad economic times, the boost in performance could be achieved by minimal modifications to existing infrastruc-

ture. A new generation of antennas appears in each complete cycle, so decision makers seem to have leveraged technology improvements.

Legacy is very important to the evolution of the Deep Space Network. The designs of all of the antennas built by the DSN can trace back to original Pioneer Station design. Antennas acquired by other means (e.g., consolidation/transfer with other networks) obviously weren't held to this standard. The impact of legacy in the DSN underscores the importance of maximizing a systems' long-term advantage while meeting current needs. Short-term thinking may give a good design up-front, but being held to that legacy can really harm the system over time. The STL system as described in Section 1.4 is a good historical example of this effect.

Technology

For a fixed architectural type (i.e., the point-to-point, ground-centric architecture of the DSN), the technology performance gain seems to level out over time. The system needs many more changes to achieve the same percentage gains in improvement as could be achieved earlier (assuming it is possible at all). There appears to be some dependency on previous changes, depending on the change subcategory. Similarly, some changes are more important than others. Can this asymptotic barrier be overcome by transitioning to a new fundamental architecture?

2.3 Commentary and Conclusions

The history of the Deep Space Network is rich with examples of the strategic evolution of systems.

The failure of the Vanguard program represents the danger of over-politicization of engineering systems while the success of Orbiter/Explorer demonstrates the benefits of well-utilized influence. The vision and legacy of Eberhardt Rechtin is proof of the power of the human factor in the success of a complex system.

A comparison of the approaches of STL and JPL to the design of a tracking network highlights several key ingredients for success:

- Choose legacy over build-new (and vice versa) judiciously.
- Clearly identify threats and opportunities and respond appropriately in a timely manner.

- Clearly identify critical path items and their associated risks and deal with them quickly and appropriately.
- Implement strategies that minimize cost and risk and gain both the short- and long-term advantages.
- Be responsive and adaptable to unexpected events.

Politics, economics, influence, growth of demand and the human factor largely drive organizational structure.

Missions were chosen to maximize scientific value (a strategy of breadth preceded one of depth).

Economic circumstances drives the cycles of acquisition and modification. Technology improvements between cycles seems to drive the implementation of new component generations during good economic times.

The antenna designs of the DSN prove that jointly maximizing long-term and short-term advantage is very important as those initial choices will strongly shape the future direction of the system.

The effect of some changes in technology depends on the changes that came before. Some types of technological changes are more important than others when it comes to the performance of the system. Increasingly marginal returns for a given architectural type appears to be the rule, but another example of the impact of an architectural shift would be required to say for certain.

Chapter 3

The Strategic Evolution of Systems

Things alter for the worse spontaneously, if they be not altered for the better designedly. - Francis Bacon

The primary objective of this thesis is to develop a framework that directly addresses legacy challenges during system operation within the context of space communication networks. A framework is "a set of assumptions, concepts, values and practices that constitutes a way of viewing reality."[49] The first challenge is that legacy influences what changes are feasible and available as well as when a change may be made and how long it may take to implement. The second challenge is that legacy influences the effectiveness and desirability of any change that is made while accounting for uncertainty. Finally, legacy influences the future constraints placed on the system as the result of making further changes.

This chapter will identify potential solutions to legacy challenges and fulfill the following objectives:

- To ascertain a set of principles based on historical examples that can guide decision makers.
- To find a method to value a physical non-market traded legacy system.
- To find a method to evaluate desirability when the benefits gained and resources expended are non-monetary in nature.
- To find a way to visualize the near- and long-term advantage.

Specific examples of the application of these principles, methods and tools will be developed in the remaining case studies.

The chapter will begin with a description of the systems of interest and discuss the principles and philosophies guiding the Strategic Evolution of Systems theory.

The majority of the chapter will develop the four elements of the framework: perspective, position, plan and pattern.

3.1 Overview

The specific systems of interest for this study are space communication networks. These networks will be described as distributed modular infrastructures and some of their key properties will be discussed.

One of the thesis objectives is to identify several key principles that decision makers should keep in mind while strategically evolving complex systems. These principles are: The Principle of Change Response and its Corrolary and the Value Principle. These principles have been developed from insights gained from the DSN case study and several other examples identified in the literature.

Finally, the Strategic Evolution of Systems framework is built upon two guiding philosophies. In a manner, these values are assumptions underpinning the methodology. The first philosophy seems to be fairly standard in modern times: operational changes should be viewed as investments in the infrastructure or system. The second philosophy, self-sufficiency, is based on common sense.

3.1.1 Systems of Interest

The primary systems of interest in this thesis are space communication networks. Communication networks can be described as distributed, modular infrastructures or systems that support the value delivery of a system-of-systems, with modular assets that enable spatial and temporal distribution of the value-generating entities. The assets are interconnected, and act in a purposeful way to support a service that delivers value to the underlying stakeholders.

Examples of such systems include large telescope arrays, adaptive mesh networks, satellite constellations (e.g., GPS, Globalstar, Iridium, and etc.), distributed satellite platforms, cellular networks, navigational systems (e.g., non-directional radio beacons) and air traffic control networks. These complex system-of-systems have inherent challenges that impede the successful management of necessary and ofttimes competitive evolutionary transitions. These difficult-to-manage attributes

include capital-intensive investments, long operational life, changing purpose and evolving service requirements.

Many of these systems are strongly limited by legacy. For example, the air traffic control network, cellular networks, non-directional radio beacons, and some instances of mesh networks have pre-existing hardware and software attributes in place. Satellite constellations may suffer from constraints imposed by legacy infrastructure.[50] Commercial legacy systems must find ways to survive changing and often fickle market forces. The evolution of distributed modular infrastructures is largely demand-driven, though not all demand is based on a commercial market. Large telescope arrays are an example where evolution is driven by changing scientific experiments and funding availability. Change in demand tends to drive a change in performance and cost of the system.

Fricke and Schulz[24] argue that the "steady insertion of new technologies is necessary" for system competitiveness. Changes are often made to save money in the long-term, to increase the quality of the product, to return to a planned schedule, to correct incomplete or flawed requirements or to meet changing requirements. Changes made for these reasons create value for the system undergoing transition.

System designers embed flexibility into new systems to adapt to future uncertainties. System operators use their experience and intuition to guide their decisions on how to transition an existing system from one state to another. However, as the background and literature review indicates, few quantitative assessments exist for evaluating the long-term consequences of exercising embedded and natural options.

Decision makers will never be able to know precisely what will happen in the future, but they can take steps to make a more informed evaluation of the potential consequences. This chapter details the Strategic Evolution of Systems framework, the proposed methodology for assessing the strategic advantage of exercising available options, including a novel attempt to quantitatively evaluate the long-term strategic position of the current and accessible architectures.

3.1.2 Principles and Philosophy

This section describes the key principles gleaned from the DSN historical case study and other sources and discusses the philosophy underpinning the theory and framework behind the Strategic Evolution of Systems.

The Strategic Evolution of Systems was defined in Chapter 1 as follows:

> *The intelligent employment of strategy to achieve the long-term advantage of a system in operation, by way of maintenance of the system as-is; modification to an existing component(s), interface(s), protocol(s), structural element(s) (legacy); retirement; or by the wholesale replacement of said system (rip-out and build new).*

Comparing the approaches of Space Technology Laboratory (STL) and Jet Propulsion Laboratory (JPL) to the design of a tracking network confirms the above definition. It also suggests a set of corresponding principles by highlighting several key ingredients for a system's success:

- Choose legacy over build-new (and vice versa) judiciously.
- Clearly identify threats and opportunities and respond appropriately in a timely manner.
- Clearly identify critical path items and their associated risks and deal with them quickly and appropriately.
- Implement strategies that minimize cost and risk and gain both the short- and long-term advantages.
- Be responsive and adaptable to unexpected events.

Principles

This thesis has identified several key principles that decision makers should keep in mind while strategically evolving complex systems. The principles have been developed from insights gained from the DSN case study and several other examples identified in the literature.

In the Deep Space Network case study, JPL was successful with the design and implementation of its tracking network, in part, because they clearly identified and quickly responded to threats and opportunities. The rehabilitation and upgrade of the 64-meter antennas (see Appendix B.4) demonstrates that the type and timing of changes to the antenna system were largely driven by pedestal degradation (a threat) and overlapping downtime for repairs (an opportunity). Threats and opportunities are events or occurrences that are likely to cause a net reduction or net increase, respectively, in the value of the system. Based on the DSN case study, this thesis proposes the Principle of Change Response and its corollary:

> **The Principle of Change Response:** *Systems undergo change in response to threats and/or opportunities.*

Otherwise, why would the system decision makers choose to change? Changes introduce risk and require an investment of precious resources.

> **Corollary:** *Changes either mitigate value reduction due to threats or enable value enhancement via opportunities.*

The Principle of Change Response and its Corollary can be further verified in the literature.

Fricke et.al.[51] identifies eight causes and rationales for change in product development: (1) needs and requirements, (2) feedback and complaints, (3) complexity, (4) degree of innovation, (5) change impacts, (6) communication and coordination, (7) time, and (8) decision discipline. Fricke and Schulz[24] identifies three more drivers of development change: a dynamic marketplace, technological evolution, and variety of environments. Each of these aspects can be extended to operational evolution, and can be categorized as a threat or an opportunity. In some cases, the causes and rationales can either be threats and/or opportunities depending on the decision maker viewpoint.

Consider that "systems to be delivered must be designed not only to meet customer or market needs, but also increasingly to meet requirements and constraints of systems sharing its operational context and throughout their entire life cycle."[24] This dynamic marketplace evolution can generate either a threat or an opportunity. Suppose a new market segment opens and a few minor changes to the infrastructure can accommodate the new requirements while still meeting the needs of the current customers. Then, the dynamic marketplace has generated an opportunity without compromising the existing system. Now suppose that the customer load had expanded to the point that the current infrastructure cannot meet the service requirements of the market segment being served. If changes are not made, customers will be lost. In this case, the dynamic marketplace has generated a threat to the system.

Clearly, other issues present themselves in operational systems, primarily failure and degradation of components. As components degrade, performance decreases, and at some point, the reduction in performance negatively impacts the overall system. When this happens, rehabilitation or repair of the component is necessary to mitigate these effects. Even replacement of the component does not guarantee the same or better performance due to different materials or manufacturing standards, or

sometimes even because the exact component has been discontinued. Furthermore, even with a new part, the remainder of the system is still aged.

On the other hand, changes can enhance the value of the system by leveraging an opportunity to increase the quality of the end-product or service. Some opportunities can enable value enhancement by saving money down the line, perhaps by increasing efficiency.[51]

Changes can be reactive or proactive. Reactive changes are those that occur after a threat has occurred, for example, changes in demand, service or requirements, as well as unexpected events such as failure of all or part of the system. Proactive changes include anticipated events such as degradation and customer demand passing a pre-set threshold.

In Crises in Network Evolution: Three Case Studies and One Proposed Solution, Etkin, Zinky and Papadopoulos[52] define a Network Crisis as "the inability of a network or its supporting organization to handle its workload because of lack of resources." One of the case studies they examine is the evolution of the Boston University campus network from broadband dumb workstations to networked personal workstations and laying fiber optic lines. Over time, many critical subsystems to the long-term vision had been implemented. "They include high-speed backbone networks, a campus conduit plan, twisted pair standards, widespread twisted pair installation, department Local Area Network (LAN), terminal servers and internetworking protocol standardization." One of the first threats to the network (or "crisis" as termed in the paper) was the lack of technical support by the vendors of the network, which had been relied upon with the first generation broadband network. Furthermore, the vendors were not prepared for a LAN with the requirements Boston University had at the time. Thus the network immediately needed major rework to run reliably. When the technical support from the vendors fell through, Boston University had to hire an in-house broadband expert to overhaul the system, costing considerably more time and money than was originally estimated. The in-house expert mitigated the effects of the vendor failure, but the value of the system was decreased from where it would have been otherwise. Other crisis issues detailed in the paper include unexpectedly high growth and a lack of human resources to handle the load, as well as the lack of fault detection and fault isolation.

The second principle proposed in this thesis is the Value Principle:

> **The Value Principle** *Unchecked threats diminish the value of the underlying assets. Ignored opportunities, while not diminishing current value, can reduce potential value.*

Returning to the Crises in Networks example, suppose the in-house expert had not been hired to overhaul the system. If this had happened, then the network would have been unable to meet even the original service requirements. As the demand increased, then the system would have been even more unprepared, further reducing the value of the network, particularly with regard to its intended objectives. Of course, had there been no response to the original crisis, would the network have been phased out due to lack of demand for the unreliable service?

The second "crisis" for the Boston University network occurred due to high growth. This happened because of a university policy to improve the science and engineering facilities. New construction and renovations of existing buildings were planned, and a decision was made to take advantage of the construction to upgrade the computing and networking resources. Thus, each of the new and renovated buildings was "pre-wired" for the broadband service. As a result, the demand for the installation exploded, forcing the second "crisis" on the Boston University network and requiring many jobs to be contracted out and a doubling of staff within IT. Hypothetically, what might have happened if the decision makers had not leveraged the opportunities provided by the construction to "pre-wire"? Either the buildings would have needed later retrofitting at additional cost, or the network might not have been as successful due to reduced demand. One could argue that the then-current value of the Boston University network would not have been diminished by ignoring the opportunity, but clearly the future value of the system would have been impacted.

Although Figure 3.1 on page 120 generically demonstrates the Value Principle, it is easy to see its application to the evolution of the Boston University network. The opportunity to "pre-wire" the buildings was exploited and for a time the value increased, but then the installation demand exceeded the resources available to IT, and the value dropped off until the wiring installations could be contracted out and the IT staff could be increased to ensure the maintenance of the network. Clearly, the actual value of the system was reduced from the initial expectation since additional expense was required for contracting out and hiring new employees.

Philosophy

The Strategic Evolution of Systems framework is built upon two guiding philosophies. In a manner, these values are assumptions underpinning the methodology. The first philosophy seems to be fairly standard in modern times. The second, although based on common sense, may be more controversial.

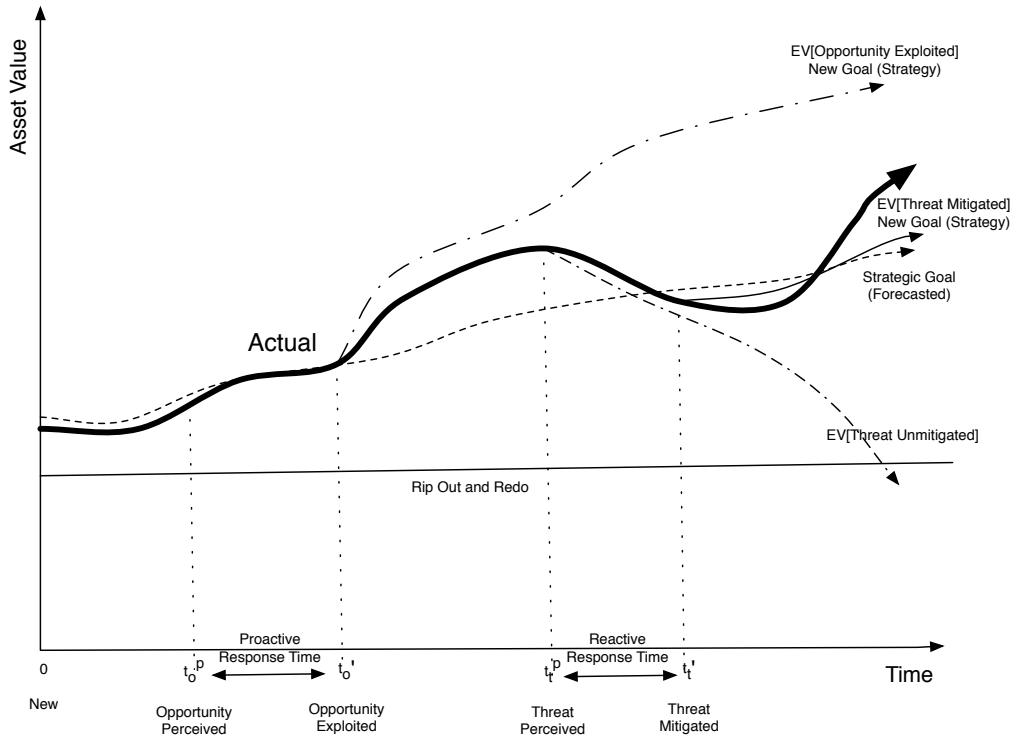


Figure 3.1: Example charting of system value over time: Note parallels to the Crises in Network Evolution Boston University network evolution example.

The first philosophy is based on the idea that operational changes should be viewed as investments in the infrastructure or system. This notion should be obvious when leveraging an opportunity is being considered. Clearly there is an investment in the system when resources are outlayed to achieve some additional objective. This same idea is less clear when mitigation of a threat is under consideration. Resources are outlayed to minimize the deleterious effects of a negative event. When recognizing that an unmitigated threat results in a system with less value, a decision maker can reduce the amount by which the value of the system will decrease. This effect can also be viewed as an investment.

Suppose the value of the system is not impacted by a threat. Then, the threat is not really a threat to the system since a threat is defined as an event or occurrence that is likely to cause a net reduction in the value of the system.

The second philosophy can be thought of as the philosophy of self-sufficiency. In other words, an infrastructure or system ideally should be able to pay for itself.

In a commercial system, this philosophy is an obvious necessity since investors demand a return on their investment, but it is much harder to obtain (or at least measure) in government systems. Are the taxes collected to cover the construction and maintenance costs of public roadways "self-sufficient"? The roadways have value, but do they pay for themselves? The net benefit of having them should justify the expense for most people, since it enables a broader range of transportation, employment and shopping venue options then if they did not exist. But the taxes levied are an indirect measurement of return.

This effect is even more true for planetary exploration systems. There is rarely a monetary return on the investment, though there are returns on the investments through commercial applications of the science learned and the technology developed for such missions. In these systems, the pay-off is more a function of the scientific value of the missions and the knowledge gained than the dollars or profits gained.

Something similar could be said for the public roadway systems. The pay-off is the mobility gained by the existence of the system. Thus, the self-sufficiency for certain systems should be thought of as a goal to improve cost-efficiency, rather than in the direct monetary return on investment.

Either way, the point at which the system is "paid off" is in the eye of the decision maker and/or stakeholders. The point of the philosophy is that the infrastructure or system should provide at least an on-par return on investment, in whatever form that takes. The application of this philosophy will become apparent in the development of the long-term position metric.

3.2 Framework

The novel framework introduced in this thesis involves the four key elements of strategy; namely, perspective, position, plan and pattern. The flowchart Figure 3.2 on page 122 and the discussion that follows shows how the definition of strategy has been expanded for use with this thesis. The remaining sections of this chapter describe each of the methodological steps shown in Figure 3.2 in detail.

To recall, strategy is the art of distributing and applying resources to fulfill the system objectives. The application of strategy to system evolution can be described with its four elements::

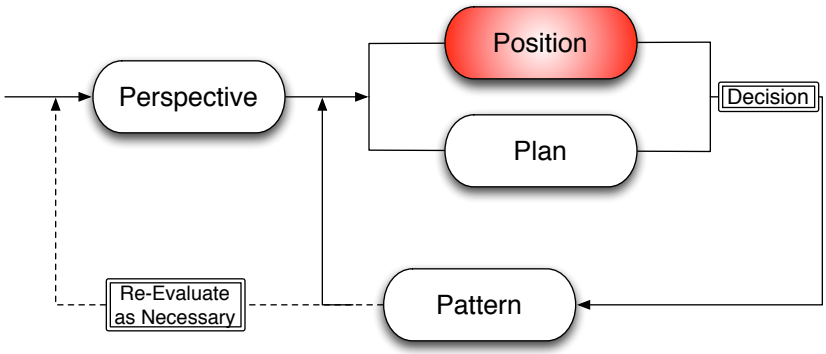


Figure 3.2: Flowchart of the Strategic Evolution of Systems framework. Position is highlighted since it is the most important and unique contribution to the framework.

- **Perspective:** Determination of the relevant information, how this information is viewed and specification of the methods used to evaluate the information.
- **Position:** Current situation (favorable or unfavorable) relative to alternatives.
- **Plan:** Guidelines for taking the position within the context of perspective and making an actionable decision.
- **Pattern:** Observations of the actual decisions by decision makers over time.

These four categories can be viewed as iterative steps, repeated as needed throughout the life cycle of the system.

In the case of a green-field system, there is not yet legacy, so the process begins with the perspective of the decision maker (though one can take the historical patterns of similar systems as a guide). Once the perspective has been identified, scoped, described and mapped, the position of the system can be quantified and qualified. A plan can be developed, and any actions taken thereafter are documented as a pattern.

In a brown-field system, the perspective has already been shaped by pattern, and one can hope it has been sufficiently documented.

Perspective: Perspective is the decision maker's worldview of the system. It is the lens by which the value of the system and any changes to it are viewed. What are the objectives of the system and how are they prioritized? What are the requirements and how should they be prioritized? What are the constraints and tensions? What are the set of decisions that differentiate the various archi-

tecture options (i.e., system configurations)? How should the architectural decisions and decision alternatives be bounded? What are the feasible set of architectural options? What changes exist to move between them and what are their limitations? What resources are required to maintain the system and to transition to each of the options? What aspects of the system generate value? What makes the system desirable and how can this be expressed? Are there options that are feasible but not available due to policy or another reason? Are there options that can be eliminated from consideration? What are the constraints for evolving the system over time (e.g., time lags for development and manufacturing)?

The aspects of perspective that need to be addressed by the decision maker are summarized in the following list. These items will be addressed in more detail in later sections:

- Identify objectives, requirements and constraints
- Identify decisions, bounds and logical constraints
- Generate relevant architecture instantiation networks
- Estimate the cost of transitioning
- Estimate the architecture desirability measured as a function of the system benefit and the resources expended to achieve those benefits
- Reduce the set of options

Position: The idea of position is the most important and unique contribution of this thesis but is difficult to define. Simplistically, position is a measure of value that defines what the system configuration enables and excludes. For position to be meaningful for a physical system, it must be measured relative to the system's alternatives. It is thus necessary to understand the structure of how the system's options relate to one another, both in terms of what changes are required to move between them and how the behavior of the system defined by each option is affected by time, uncertainty and the path taken through this option space.

Position is meant to be viewed using the decision maker's perspective. If evaluated accurately, position should provide the decision maker with insight into the relative behavior of the system's options, thus indicating dominant desirable configurations. This dominance should help the decision maker drive toward systems that advantageously trade the objectives of maximizing benefits and opportunity and minimizing costs and risk. Therefore, position should be specifically defined with this goal in mind. (Note: The position of a system is only as meaningful as the information used

to evaluate it. There is no guarantee that the position that is evaluated using this method will be accurate — only that its use provides the decision maker with structural information about the options before him that may not be available in any other way.)

This thesis proposes a novel approach for evaluating an existing system's position relative to its alternatives. This approach represents a preliminary formulation for capturing the complexities of position and is by no means the only way that position might be defined. As will be shown later in this chapter and in the subsequent case studies, this approach appears to work very well but suffers from a sensitivity to the chosen horizon time and to the choice of factors meant to reduce the effect of certain biases inherent in the formulation.

Several of the steps (e.g., the time horizon, the weightings and debiasing factors, the terminal entropy, and the propagation of energy and entropy) require explanation of the core elements of the approach before they can be discussed. For this reason, the core elements, such as the reason for using energy and entropy and the derivation of the entropy formulation will be discussed later in this chapter. The other steps will be addressed where appropriate and the overall approach will be demonstrated in the subsequent case studies.

The approach to position is outlined here:

- Identify an appropriate time horizon
- Identify appropriate weightings and debiasing factors
- Describe the system desirability at an instant in time using energy
- Model the transitions between architectural options using energy accounting
- Propagate the energy forward in time until the time horizon is reached (evaluate over many scenarios)
- Estimate the terminal entropy values at the time horizon
- Propagate the position entropy representing the system's desirability (dynamic, multi-dimensional value) from the time horizon back towards the current time
- Evaluate the difference in the final position entropy between the legacy system and alternatives (transition entropy)
- Plot these transition entropy values (multiple scenarios create regions corresponding to each alternative)

Plan: Plan behaves as a roadmap decision-making process. The decision maker should be able to use the position information as viewed from his perspective and develop a strategic course of action. It may be clear from the position information that maintaining the current system for another period of time in order to wait for some uncertainty to resolve may make the most sense. The position information may also indicate that an immediate transition to another configuration with low-level preparations for a future transition is the most strategic plan. In other words, using position and perspective to build a plan enables the decision maker to put together a series of actions with an indication of the approximate level of resources to expend in the next time period.

Pattern: Pattern serves to provide continuity, encourage documentation and establish a standard method for evolving the system. These often overlooked aspects of system evolution are especially important for long-life systems.

3.3 Perspective

Perspective is the decision maker's worldview of the system. It is the lens by which the value of the system and any changes to it are viewed. It can best be defined by a series of questions: What are the objectives of the system and how are they prioritized? What are the requirements and how should they be prioritized? What are the constraints and tensions? What are the set of decisions that differentiate the various architecture options (i.e., system configurations)? How should the architectural decisions and decision alternatives be bounded? What are the feasible set of architectural options? What changes exist to move between them and what are their limitations? What aspects of the system generate value? What makes the system desirable and how can this be expressed? What resources are required to maintain the system and to transition to each of the options? Are there options that are feasible but not available due to policy or another reason? Are there options that can be eliminated from consideration? What are the constraints for evolving the system over time (e.g., time lags for development and manufacturing)?

The aspects of perspective that need to be addressed by the decision maker are summarized in the following list. The first step comprising the perspective of the decision maker is largely based on standard industry practice. The others build upon previous research. These items will be addressed in detail in this section.

- Identify objectives, requirements and constraints
- Identify decisions, bounds and logical constraints
- Generate relevant architecture instantiation networks
- Estimate the cost of transitioning
- Estimate the architecture desirability measured as a function of the system benefit and the resources expended to achieve those benefits
- Reduce the set of options

3.3.1 Identify objectives, requirements and constraints

The first step in describing the perspective of the decision maker is to identify the objectives, requirements and constraints of the system. There are well-published industry standards for system-level design. Example publications of industry practice include the International Council on Systems Engineering (INCOSE) standards in its Systems Engineering handbook[53] as well as the process descriptions in Space Mission Analysis and Design (SMAD)[54]. SMAD provides several key definitions, including:

Objectives: Broad goals which the system must achieve to be productive.

Functional requirements: Define how well the system must perform to meet its objectives.

Operational requirements: Determine how the system operates and how users interact with it to achieve its broad objectives.

Constraints: Limit the cost, schedule and implementation techniques available to the system designer.

The Strategic Evolution of Systems framework extends the scope of this step (identifying objectives, requirements and constraints) by including the identification of evolution-level objectives, requirements and constraints. These include:

- Projected (i.e., likely and/or desired) objectives and requirements over time.
- Specification of time horizon(s).
- Reductive time-dependent constraints on feasible decisions, such as the identification of any time lags between making a decision and its implementation.

3.3.2 Identify decisions, bounds and logical constraints

The second step to describe the perspective of the decision maker is based on Simmons' Architecture Decision Graph (ADG) methodology.[55] In particular, the relevant steps are:

- Bound the architecture space.
- Find the set of decisions that "potentially changes the overall high-level concept of the architecture to be implemented."
- Identify the logical constraints between the decisions and decision alternatives.

These steps are best illustrated by describing their use in the construction of the Lunar Outpost Architecture Study, a case study performed by Simmons in his doctoral thesis.

Bounding the architecture space

The Lunar Outpost Architecture study initially bounds the space according to policy statements and congressional mandates. The decision variables are assumed to be bounded based on the options enumerated in the "Lunar Architecture Update" document,[56] although the case study adds the possibility of an intermediate outpost. Furthermore, the case study is limited to addressing whether certain lunar exploration campaign elements should be included or excluded from further study.

Generically, bounding the architecture space means limiting the number and type of decisions and decision alternatives to ones that are meaningful to the case at hand. For example, one should limit the architecture space to those decision alternatives that are mature and readily available for an initial look at the space of outcomes. The value of investing in immature technologies may become apparent during successive iterations of the methodology by relaxing these bounds.

Finding architecturally distinguishing decisions

The Architecture Decision Graph methodology defines architecturally distinguishing decisions as those that impact the mapping of architectural functions to architectural forms. In the case of the Lunar Outpost Architecture study, the architecturally distinguishing decisions include campaign elements based on the high-level lunar exploration campaign models described in Hofstetter, et.al.,[57] as well as the set of options provided in the NASA Lunar Update.[56] The inclusion of Exploration

Systems Architecture Study (ESAS)-style long sorties and intermediate outposts as well as the set of their potential locations are examples of the campaign-element decisions incorporated into the case study. Other decisions included key questions surrounding the transportation architecture, surface mobility elements, human habitation, outpost power and communication sub-architectures. Several simplifying assumptions were used to reduce the set of possible decisions, including the assumption that there is always a long-term outpost.

Simmons specifies the architecturally distinguishing decision in a morphological decision table that includes identifying the decision (e.g., outpost energy storage options), the available alternatives (e.g., none, batteries, fuelcell), and the units of the alternatives (e.g., none).

Identifying logical constraints

Simmons describes logical constraints as constraints on "the set of feasible combination of assignments to decision variables." An example of a logical constraint is that if pressurized connections are available between the habitat and rover, then the rover must be pressurized in order to accommodate the connections. Simmons assigns a name, defines the scope of the logical constraint (the decision variables affected), and specifies the contraint using logical operators.

The Strategic Evolution of Systems extends this definition to include logical constraints on the feasible decisions at the evolution level, as certain constraints may change depending on the level of demand. For example, it may not make sense to transition a system that can command one level of usage fees based on its performance characteristics to one that generates a lower level of usage fee.

3.3.3 Generate the Architecture Instantiation Networks

The steps to generate the relevant architecture instantiation networks are based on the ideas and methods proposed in various works.[58,55,15,16].

An Architecture Decision graph is "an atemporal representation of a decision problem so that the decision ordering can be automated through analysis of the structure of the problem, rather than requiring the order of the decision variables to be pre-specified."[55]

The method described by Ross and Hastings plots a Utility vs. cost trade space and applies specified transition rules to form a trade-space network. The proposed metric is the filtered outdegree, which

is the number of outgoing arcs from a design, filtered by some acceptable "cost."[58] A primary limitation of the changeability method is that it only looks one transition out, which is fine if it is expected that only one operational change will be made.

The Time-expanded Decision Network method proposed by Silver uses very similar steps to Strategic Evolution, although the methodology does not specifically enumerate transition rules other than to say that the switches between configurations must account for the costs of all possible switches.[15] Furthermore, it is assumed that the family or set of designs has already been found. The method is similar to the method proposed here with the several distinctions. First, it is intended for use in aiding in initial design in order to improve the future evolvability of a system, and it focuses on minimizing life cycle costs. The complex systems focused on are heavy lift vehicles, which are individual systems within the context of a larger system (space exploration). This thesis extends the idea to distributed systems. We shall see how the Strategic Evolution of Systems builds on these ideas and relaxes the inherent constraints of Silver's Time-expanded Decision Network methodology.

The steps for generating the relevant architecture instantiation networks for the Strategic Evolution of Systems methodology are as follows:

- Identify transition rules.
- Identify architecture families.
- Identify current architecture family and state (if it exists).
- Apply transition rules to current architecture family.
- Identify supernodes.

These steps are discussed in detail in the subsections below.

Identify Transition Rules

Ross and Hastings define transition rules as "mechanisms" to get from A to B. Each rule specifies the design variables that must be changed in order to make the transition, whether it's reversible or irreversible (including direction of irreversibility), and the path enablers required, if any. Path enablers are defined as "intervening parameters that reduce the cost for transition paths for a design, including creation of the path option itself." An example of a path enabler is the use of a space tug for changing orbital planes.[58]

Simmons provides a method for arriving at key decisions, while Ross and Hastings give a way to take those decisions and find feasible transitions between them. A limitation of Simmons' Architecture Decision Graph method is that it is designed to provide a comprehensive view of the decision space in order to prioritize and order decisions during the conception and design decision development phase and to map out the impact of changes to those decisions relative to downstream decisions, and is not intended for use in making evolutionary design and operational decisions.

There are several types of transition rules:

- "Infeasible" to change in all cases (no arc).
- "Infeasible" to change in some cases, also known as an irreversible change, such that once a change is made, it cannot return to the previous architecture (directed arc).
- "Feasible" to change in all cases, also known as a reversible change (undirected arc).

Identify Architecture Families

Architecture families arise when it is infeasible to change one or more decision variables in either direction. The system can be reconfigured to any of the accessible architecture instantiations according to the transition rules established for the adjustable decisions. However, once a specific decision assignment has been made to one of the fixed-decision variables, the system cannot be moved to another architecture family without ripping out the existing system and building new. In a sense, this is architectural "lock-in" since the system is locked into the set of configurations accessible within the current family.

Identifying the current architecture family and state The primary difference between the research in the literature and the Strategic Evolution of Systems is that it is assumed here that there is an existing system in operation. Thus, the decision makers should be able to find the set of decision assignments that specify the current state. Doing so will also identify the current architectural family, so the space of potential architectures can be bounded to the current family. The decision on which architectures to consider for the ripping out and building new option is an entirely separate art form and is largely left to future research. The rip out and build new option is considered in this thesis but has been bounded to fall within the current architecture family.

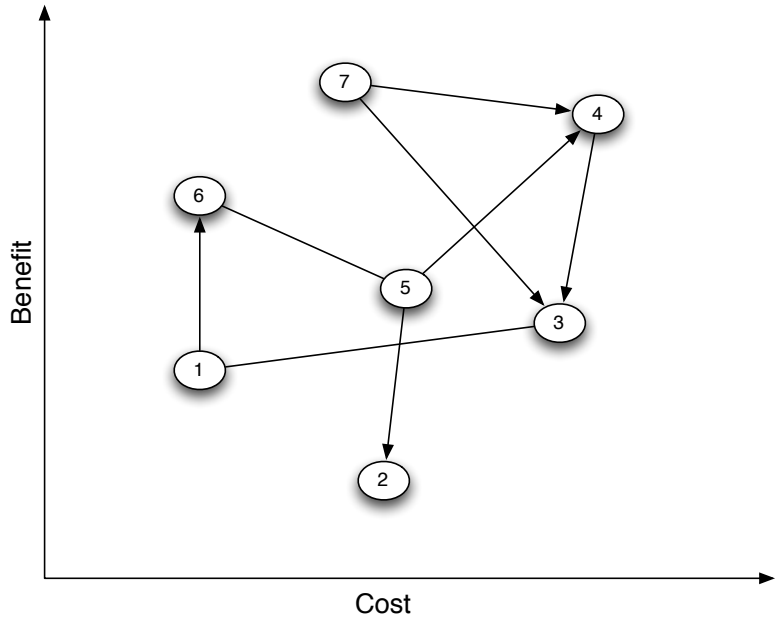


Figure 3.3: Example architecture instantiation network where nodes represent system configurations and arcs represent the possible transitions between them. Transitions can be reversible, irreversible or nonexistent. Here, the configurations are plotted on a static benefit-cost plane.

Generate Architecture Instantiation Network

Now that the transition rules have been identified and the set of designs has been constrained to the current architecture family, the rules can be applied to the reduced set to generate the architecture instantiation network. This network will be used to evaluate the value of the legacy architecture and to aid in identifying the most advantageous transitions to make as time progresses. An example of such a network is shown in Figure 3.3. The nodes of the architecture instantiation network represent the feasible and available system configurations and the arcs represent the possible transitions between them. The transitions may be filtered by acceptable cost as proposed by Ross and Hastings.[58]

Identify Supernodes

Supernodes are defined in this thesis as the set of architectures that are delimited by irreversible transition rules. Within a supernode, all architectures are fully reversibly connected. The state of the system can freely migrate around these architectures, but there is a cost penalty associated with transitioning between them. Migration from supernode A to supernode B automatically means that

the set of architectures in supernode A are no longer accessible unless some currently unknown method becomes available to generate a new transition rule opening that space of architectures back up.

Supernodes are interesting for several reasons. It should be clear that some supernodes act as sources (from within the supernode, all other architectures are accessible), some act as relays (the only way from supernode A to supernode B is to go through supernode C), and others act as sinks (architectures can migrate from other supernodes into the sink supernode, but once in the sink supernode, there is no foreseeable escape).

Of greatest interest is that by clustering architectures within supernodes, the underlying relationship structure between the architectures is illuminated.

3.3.4 Estimate the Cost of Transitioning

Once the mechanisms for transitioning have been identified, their costs can be estimated. Costs can include hardware and software development, manufacturing/implementation and testing; the costs to employ human resources; and policy costs (e.g., paperwork). However, costs can also be associated with the downtime necessary to transition the system, the time it takes the transition, and the risk inherent to any system change (e.g., transition incomplete, performance lower than expected, and etc).

3.3.5 Estimate the Architecture Desirability

A significant portion of this thesis is devoted to developing the theory behind estimating the desirability of an architecture, the use of this desirability in evaluating the set of potential transition options, and the mapping of the consequences of said changes to the desirability of an architecture during a given time horizon. The architectures and their performance must be modeled and simulated during this step. It is important to identify what makes an architecture desirable and how this desirability is affected by internal and external drivers. These drivers must be identified and the uncertainties inherent in their movement over time modeled. The costs of operating the system must be estimated. The art of strategy comes into play here very strongly.

Value is a static measure of benefit at cost, taken at a snapshot in time, with a specified resolution to uncertainty and at a fixed location in the option tree. Desirability is also a function of the system

benefits and the resources expended to obtain those benefits, but is the dynamic multi-dimensional value of a system, an observation of the value of the system over time, uncertainty, and the path taken through the option space. A system is desirable if it has value and it has long-term, robust value if it is desirable.

There appears to be four main components of desirability: to maximize the benefit gained, minimize the resources expended (cost), maximize the opportunities enabled, and minimize the risks — potential threats — undertaken. Although benefits and costs vary over time, their accrual is only meaningful when considered at a fixed point in time (e.g., present value, future value). Static system desirability is defined as a function of the current benefit acrrual and the current resource expenditure.

In contrast, opportunities and risks are more dynamic. Opportunities and risks are not fixed in time: they are future events that may or may not happen and further action is required to enable an opportunity or to mitigate a risk. Furthermore, some opportunities and risks are dependent on the structure of options that may be exercised (i.e., transitioning to architecture A enhances benefits and reduces the required resources for a certain set of conditions, but if those conditions are no longer met, architecture A is less desirable than the current architecture and perhaps it is no longer possible to return to the current architecture and the transitions that can be made are not very desirable, either).

The form of the static desirability of an architecture will need to be evaluated by the decision maker. It is proposed here that this desirability will be some function of the benefits gained and the resources expended. Which benefits and what resources are up to the decision maker. The form of the equation is up to the decision maker as well, although a specific metric to use is proposed in the Position section.

The dynamic nature of desirability is dependent on the system response to uncertain conditions, the structure of options that may be exercised, and the path taken through this option space (legacy). These dependencies are complex and difficult to model. In order to estimate the dynamic desirability of an architecture, the system response to various conditions must be modeled. The conditions to be evaluated must be carefully identified and bounded. The relationships between the various architectures need to be found using the architecture instantiation network method described earlier.

This thesis proposes the use of an energy analogy for the modeling of the static system desirability and a unique entropy formulation for modeling the dynamic system desirability. These approaches

will be discussed in detail in the Position section of this chapter.

3.3.6 Reduce the Options

If the number of feasible options is large, then it might be desirable to reduce the set of options. Feasible options include architectures that are technologically possible and meet the requirements and constraints. As will be shown in the case studies, more than seven options can be difficult to evaluate within a short period of time. The case studies will demonstrate numerous methods to reduce the set of options. Several of these will be discussed here for illumination.

There are times when options may be feasible but are not available due to policy reasons. For example, small nuclear reactors are great for powering spacecraft, but their use is restricted due to environmental reasons. Options may become unavailable for other reasons. Obsolescence may reduce or eliminate the required manufacturing facilities for certain technologies.

Sometimes there are options in the instantiation network that perform worse than the current architecture. The case studies demonstrate a type of reduction to eliminate options that do not make sense from a static performance standpoint.

Option reduction is an important artform — it is undesirable to trim away options that end up being desirable later!

3.4 Position

The idea of position is the most important and unique contribution of this thesis but is difficult to define. Simplistically, position is a measure of value that defines what the system configuration enables and excludes. For position to be meaningful for a physical system, it must be measured relative to the system's alternatives. It is thus necessary to understand the structure of how the system's options relate to one another, both in terms of what changes are required to move between them and how the behavior of the system defined by each option is affected by time, uncertainty and the path taken through this option space.

Position is meant to be viewed using the decision maker's perspective. If evaluated accurately, position should provide the decision maker with insight into the relative behavior of the system's options, thus indicating dominant desirable configurations. This dominance should help the decision

maker drive toward systems that advantageously trade the objectives of maximizing benefits and opportunity and minimizing costs and risk. Therefore, position should be specifically defined with this goal in mind. (Note: The position of a system is only as meaningful as the information used to evaluate it. There is no guarantee that the position that is evaluated using this method will be accurate — only that its use provides the decision maker with structural information about the options before him that may not be available in any other way.)

This thesis proposes a novel approach for evaluating an existing system's position relative to its alternatives. This approach represents a preliminary formulation for capturing the complexities of position and is by no means the only way that position might be defined. As will be shown later in this chapter and in the subsequent case studies, this approach appears to work very well but suffers from a sensitivity to the chosen horizon time and to the choice of factors meant to reduce the effect of certain biases inherent in the formulation. The approach to position is outlined here:

- Identify an appropriate time horizon
- Identify appropriate weightings and debiasing factors
- Describe the system desirability at an instant in time using energy
- Model the transitions between architectural options using energy accounting
- Propagate the energy forward in time until the time horizon is reached (evaluate over many scenarios)
- Estimate the terminal entropy values at the time horizon
- Propagate the position entropy representing the system's desirability (dynamic, multi-dimensional value) from the time horizon back towards the current time
- Evaluate the difference in the final position entropy between the legacy system and alternatives (transition entropy)
- Plot these transition entropy values (multiple scenarios create regions corresponding to each alternative)

Several of the steps (e.g., the time horizon, the weightings and debiasing factors, the terminal entropy, and the propagation of energy and entropy) require explanation of the core elements of the approach before they can be discussed. For this reason, this section will focus on the core elements, such as the reason for using energy and entropy and the derivation of the entropy formulation. The other steps will be addressed where appropriate and the overall approach will be demonstrated in

the subsequent case studies.

3.4.1 Energy Analogy to Model System Desirability

The idea for using an energy analogy to model the static system desirability developed from a realization that investing resources in a system in order to move to a more desirable architecture was analogous to adding energy to a chemical substance in order to create a new substance. Just as a chemist deliberately adds a certain amount of energy to hydrogen and oxygen in order to arrive at water, which has more desirable properties for certain purposes, decision makers choose to invest resources in order to achieve a system with more desirable properties.

In chemistry, this process is known as activation energy (see Figure 3.4 for hydrogen and oxygen becoming water). Activation energies act as hurdles to prevent substances from spontaneously changing.[2]

Verification of this analogy is found in a remarkably similar diagram in Figure 3.5 illustrating the effects of the transition from IPv4 to IPv6.[59]

It should be evident that to apply the analogy, less energy must imply greater desirability. There are a number of ways to define energy, depending on the type of system in question. The metric used in this thesis is described in the following section and the reasoning for its derivation should provide a template for how to choose a system-appropriate definition.

A simplistic definition of energy as applied to engineered systems can be gleaned from the IPv4 to IPv6 diagram. In this case, energy equals cost. However, this definition ignores the other side of the equation: benefit. A system could cost very little but have very little benefit, whereas a system costing more could have a significantly greater benefit. There needs to be consideration of both cost and benefit.

In this thesis, energy is defined as a function of cost and benefit. For the purposes of illustration, let energy be defined over **t** discrete time periods as shown in Equation 3.1:

$$E(t) = \frac{\sum_{i=1}^{t} cost(i)}{\sum_{i=1}^{t} benefit(i)} \tag{3.1}$$

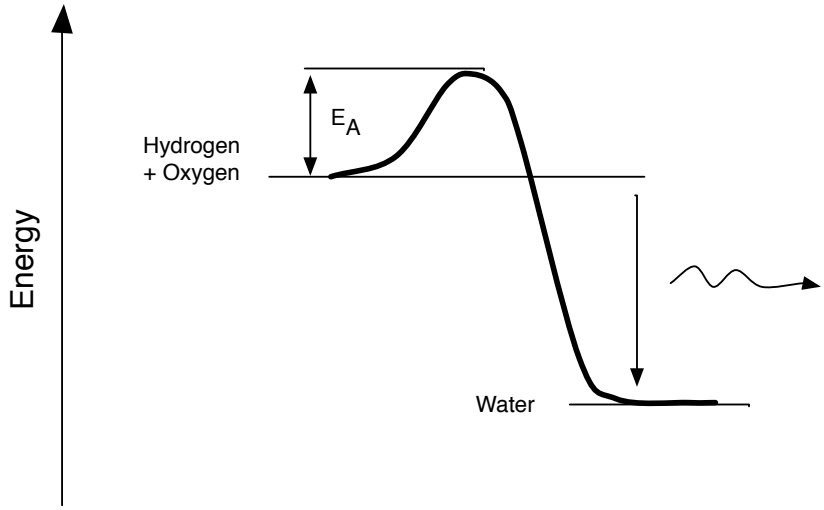


Figure 3.4: Activation energies act as hurdles to prevent substances from spontaneously changing. In this example, an activation energy prevents hydrogen and oxygen from spontaneously converting to water. Figure adapted from Lambert.[2]

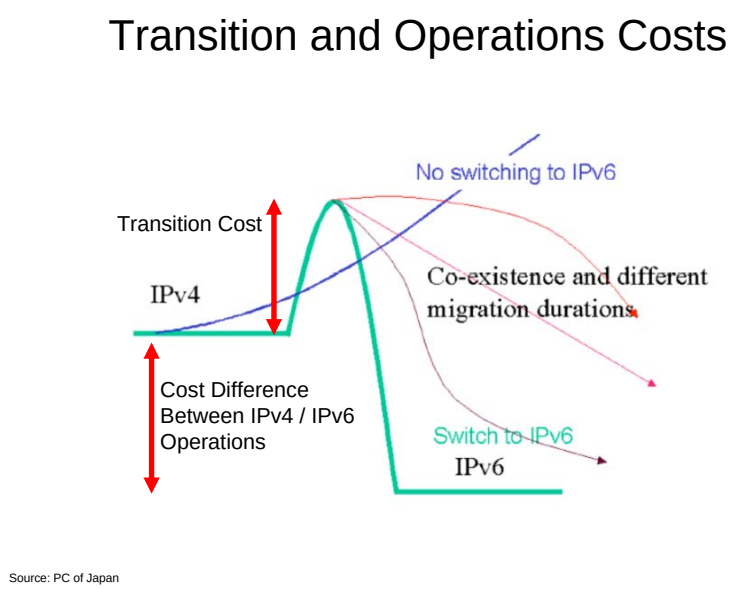


Figure 3.5: The cost effect of transitioning from IPv4 to IPv6. Picture taken from Ivancic.[3]

Using discrete time periods simplifies calculations and enables modeling of commercial systems with revenues coming in over predictable periods (e.g., monthly usage fees). For non-commercial systems, the use of discrete time periods enables modeling of operation stages. It would be a trivial matter to convert to continuous time; just replace the summation with an integral.

This definition is essentially an inverse of the benefit-cost (also known as cost-benefit) ratio metric so popular in economic evaluations of large-scale public projects. Because of this relationship, it is fairly easy to project the properties inherent to this definition of energy. It is well known that a benefit-cost ratio of greater than 1.0 is preferable since the cost investment of the project is less than the benefits gained by implementing the project. The greater the value of this ratio, the more preferable the project. This property of the ratio implies the following advantages to the metric. Projects can be compared on a common scale and they can be ranked on the basis of their preferability, which gives a clear indication as to whether the project is worthwhile (value relative to 1). Disadvantages include the obvious bias against projects with large recurring (operational) costs and toward capital-intensive projects, the requirement that all benefits be described in terms of monetary value, and the ambiguity with how to treat recurring costs.[60]

The benefit-cost metric has traditionally been used to evaluate "build/no build" decisions with a set of alternative green-field projects. It is generally avoided by organizations involved in projects with large recurring costs, and is almost never used in business.[60]

At first glance, this would appear to be a poor choice of definition for energy. However, this thesis is focused on distributed modular infrastructures, in particular communication networks. In nearly all cases, these types of systems fall under the category of capital intensive projects, which are the types of projects that tend to use the benefit-cost metric.

It is proposed that the disadvantage associated with the need to define benefits in terms of monetary value can be mitigated. Since unity energy implies that the cost of the project has been paid off, the non-monetary benefits can be scaled to reflect the decision-maker's view of return on investment. Utility theory can be used to find the preference curve of the decision maker, scaled to ensure that unity energy retains its property.

An energy greater than unity implies that the total benefits gained have not yet caught up with the total cost invested, while an energy less than unity implies that the total benefits gained are greater than the total cost invested. The goal is to attain as much bang for the buck, so to speak.

The ambiguity of benefit-cost analysis with regard to how to treat recurring costs can be mitigated

to some extent as well. Since the benefits are not necessarily monetary in nature, it no longer makes sense to consider the ratio of net benefits over initial investment. The additional treatment of the transition costs complicates things a bit. It is proposed that the transition costs be considered as separate capital expenses. The transition costs are one-time expenses, which make them capital expenses. Every option under consideration shares the same history, so the initial capital expense will be the same. The initial investment expense is known as it is in the past, but the future transition costs are uncertain, so keeping them separate enables evaluation of the risks. It is also possible to use the transition cost contribution to evaluate the potential value of finding ways to reduce the transition costs.

The following energy accounting section goes into the types of transitions and the various contributions in more detail.

3.4.2 Energy Accounting to Model Transitions

Energy accounting is a method to account for the effect of various types of options on a system's energy. There seem to be four main classes of options:

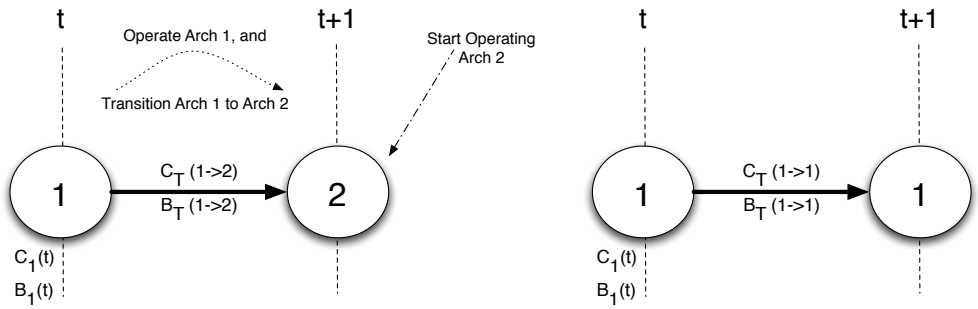
- Transition from one architecture to another within a family.
- Maintain the current architecture "as is."
- Rip-out the current architecture and build a new architecture (either within the same family or within a new family).
- Retire the current architecture without replacement.

In each case, the previous values for the cumulative benefit and cumulative costs must be estimated, as well as the increases in both over the period of interest.

Each of the contributions described below can be tracked. Variations on the expected values can be plugged in to map out the potential downstream consequences of any differences between expectation and reality.

There are two assumptions made for all cases: first, it is assumed that the time to transition between the architectures takes the entire period; and second, the legacy system — Architecture 1 — continues operating while the transition is occurring.

The following sections discuss the proposed energy accounting for the four types of options.



(a) Transition (b) Maintain As-is

Figure 3.6: Energy accounting for the (a) transition and (b) maintain "as is" cases.

Transition between Architectures

There are several things that happen when Architecture 1 is transitioned to a new Architecture 2. There is the cost to transition between the architectures over the period, the operating costs over the period and the benefits associated with operating the architecture over the period given the transition is occurring. These changes are shown in Figure 3.6a on page 140 and captured in Equation 3.3.

$$E_1(t) = \frac{C_1(t)}{B_1(t)} = \frac{\sum_{i=1}^{t} C(i)}{\sum_{i=1}^{t} B(i)} \tag{3.2}$$

$$E_2(t+1) = \frac{C_2(t+1)}{B_2(t+1)} = \frac{C_1(t) + C_T(1 \rightarrow 2)}{B_1(t) + B_T(1 \rightarrow 2)} \tag{3.3}$$

where:

- $C_T(1 \rightarrow 2)$ is the cost of operating Architecture 1 over the time period AND the cost of transitioning between Architectures 1 and 2.
- $B_T(1 \rightarrow 2)$ is the benefit of operating Architecture 1 over the time period GIVEN the architecture is being transitioned from Architecture 1 to 2.

Maintain the Current Architecture

When the current architecture is maintained, the only adjustment to the cost accumulation is the maintenance (and repair if necessary) costs over the period. The additional benefits gained by operating over the period must also be included in the accounting. These changes are shown in Figure 3.6b on page 140 and captured in Equation 3.5.

$$E_1(t) = \frac{C_1(t)}{B_1(t)} = \frac{\sum_{i=1}^{t} C(i)}{\sum_{i=1}^{t} B(i)} \tag{3.4}$$

$$E_1(t+1) = \frac{C_1(t+1)}{B_1(t+1)} = \frac{C_1(t) + C_T(1 \rightarrow 1)}{B_1(t) + B_T(1 \rightarrow 1)} \tag{3.5}$$

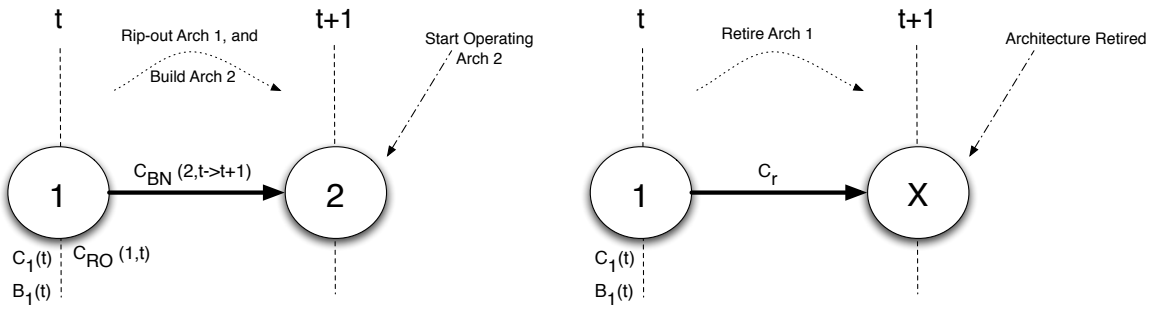
where:

- $C_1(t), B_1(t)$ are the costs incurred and benefits gained, respectively, for Architecture 1 until time t.
- $C_T(1 \rightarrow 1), B_T(1 \rightarrow 1)$ are the costs incurred and benefits gained, respectively, over the operating period for Architecture 1.

Rip Out and Build New

There are two types of costs incurred in this class of options. The first set of costs is associated with the cost of ripping out the current architecture; the second set with the building of the new architecture. Clearly, there are no operating costs because during this period, there is no architecture in place. The same goes for the benefits; no operating architecture means no benefits are gained by operating the system. These changes are shown in Figure 3.7a on page 142 and captured in Equation 3.7.

$$E_1(t) = \frac{C_1(t)}{B_1(t)} = \frac{\sum_{i=1}^{t} C(i)}{\sum_{i=1}^{t} B(i)} \tag{3.6}$$



(a) Rip-out and Build New (b) Retire

Figure 3.7: Energy accounting for the (a) rip out and build new and (b) retire cases.

$$E_2(t+1) = \frac{C_2(t+1)}{B_2(t+1)} = \frac{C_1(t) + C_{RO}(1,t) + C_{BN}(2,t+1)}{B_1(t)} \tag{3.7}$$

where:

- $C_1(t), B_1(t)$ are the costs incurred and benefits gained, respectively, for Architecture 1 until time t.
- $C_{RO}(1,t), C_{BN}(2,t+1)$ are the costs incurred from ripping out Architecture 1 at time t, and building Architecture 2 during the period (t,t+1], respectively.
- There are no benefits gained during the period (t,t+1] since there is no architecture in place for operation.

Retire

When an architecture is retired, there are costs associated with its obsolescence. In some cases, this means the costs associated with deactivating and deorbiting satellites. In other cases, it might mean the costs associated with ripping out train tracks. It may also means the cost of transferring ownership to another entity. For the purposes of this thesis, it is assumed that there is no benefit derived by retiring the system. These changes are shown in Figure 3.7b on page 142 and captured in Equation 3.9.

$$E_1(t) = \frac{C_1(t)}{B_1(t)} = \frac{\sum_{i=1}^{t} C(i)}{\sum_{i=1}^{t} B(i)} \tag{3.8}$$

$$E_r(t+1) = \frac{C_r(t+1)}{B_r(t+1)} = \frac{C_1(t) + C_r}{B_1(t)} \tag{3.9}$$

where:

- $C_1(t), B_1(t)$ are the costs incurred and benefits gained, respectively, for Architecture 1 until time t.
- C_r is the cost incurred by retiring Architecture 1 over the period (t,t+1].
- There are no benefits gained during the period (t,t+1] since there is no architecture in place for operation.

3.4.3 Entropy as a Measure of Position

This thesis uses entropy to measure the dynamic system desirability, which is dependent on the system response to uncertain conditions, the structure of options that may be exercised and the path taken through this option space (legacy). This dynamic system desirability must be captured and evaluated at the current time. It is for this reason that the thesis has introduced the concept of position.

Position is a measure of value that defines what the system configuration enables and excludes. For position to be meaningful for a physical system, it must be measured relative to the system's alternatives. It is thus necessary to understand the structure of how the system's options relate to one another in terms of what changes are required to move between them and how the behavior of the system defined by each option is affected by time, uncertainty and the path taken through this option space.

To conceptually understand the meaning of "position," visualize a chess game such as the one in Figure 3.8 on page 144. Please note the author is not a master of chess and uses the game only to illustrate the concept of position.

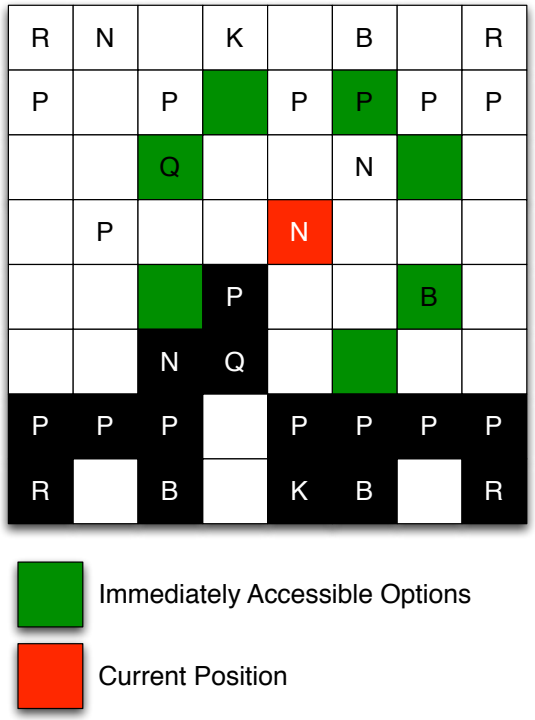


Figure 3.8: Example chess game to demonstrate the conceptual meaning of position. For reference: R (rook), P (pawn), N (knight), Q (Queen), B (Bishop). Figure adapted from chess game screen shot[4]

If the current architecture is the knight highlighted in red, the immediate options for transitioning the system are highlighted in green. Each option has a very different consequence in terms of the future position of the knight. For example, the knight can take the queen which would put the opponent in check, but the knight would immediately be taken by the opposing knight. So although the immediate change has a high value (take knight), there is a long-term negative consequence of doing so. This option is not very desirable in the long-term.

However, if the knight takes the pawn, something much more interesting happens. The immediate outcome is not very desirable, but the long-term view shows a very desirable outcome. By moving to the position of the pawn, either the king or the rook could be taken in the next move without any immediate danger from opposing forces, thus ensuring a check (at least for a short time).

Some of the value of taking the pawn is in the number of high-value options that are opened up by making the transition.

The chess analogy enables this thesis to articulate a definition for long-term position; namely, a

measure of a given architecture's access to future architectures, in terms of both the number and desirability of those options.

In other words, the more future options there are and the better they are, the greater the value of the long-term position of the current state.

But how can this definition be formulated mathematically?

First, step back a moment and consider the simplest aspect of desirability; the more options there are, the greater the value of the measurement.

This description should be familiar to those with some background in chemistry and physics.

Recall the statistical thermodynamics definition of entropy: The more microstates there are that correspond to the observed thermodynamic macrostate, the greater the value of the entropy. An increase in the number of accessible microstates implies an increase in entropy.[61] Thus, the more accessible states (i.e., options) there are, the greater the entropy.

The challenge with this formulation is the need to ensure that the better the options, the greater the entropy.

An interesting aspect of using entropy is that it is well known that a system described by entropy will evolve to the state with the most homogeneous probability distribution.[62] This tendency must be noted.

Energy and Entropy

The formulation so far assumes a relationship between energy and entropy. To verify this relationship, consider the impact of molecular transitions on state enthalpy ΔH as demonstrated in Figure 3.9 on page 146.

The state enthalpy is related to the Gibbs entropy according to the Gibbs Free energy, Equation 3.10.

$$\Delta G = \Delta H - T\Delta S \tag{3.10}$$

where the Gibbs entropy is given by Equation 3.11.

$$S = -k_B \sum p_i ln(p_i) \tag{3.11}$$

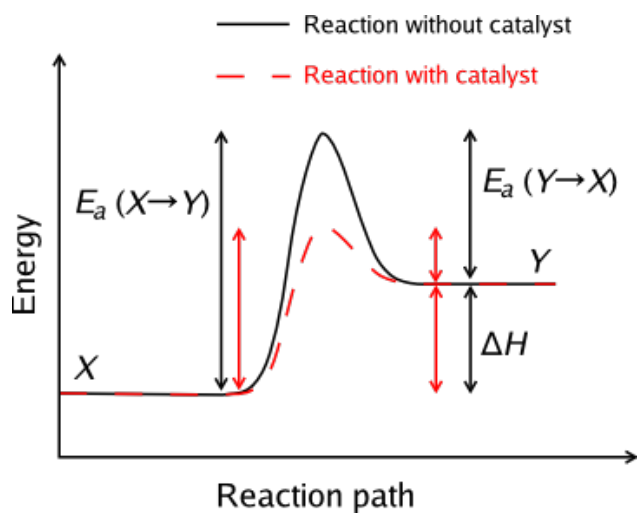


Figure 3.9: Relationship between state enthalpy and energy during a molecular transition. Figure taken from Wikipedia.[5]

such that p_i is the probability that particle i will be in a given microstate.

If the Gibbs Free energy equation is divided by the temperature T, an expression for the change of entropy results.

This relationship is important because the state enthalpy is directly related to the activation energy of the substances in question. Thus, there is a measurable physical relationship between energy and entropy on the basis of the same analogy used to relate energy and desirability in a previous section.

There are two entropy definitions of distinct interest: thermodynamic entropy and configuration entropy. Each definition has a different interpretation of the generic entropy formulation in Equation 3.12.

$$S = k_B \ln \Omega \tag{3.12}$$

where k_B is Boltzmann's constant.

Thermodynamic Entropy Thermodynamic entropy has been defined as "a measure of the energy dispersal for a system by the number of accessible microstates, the number of arrangements (each containing the total system energy) for which the molecules' quantized energy can be distributed, and in one of which — at a given instant — the system exists prior to changing to another."[63]

This interpretation assumes that the probability omega is the number of microstates corresponding to the observed thermodynamic macrostate.

Configuration Entropy Configuration entropy has been defined as "the entropy associated with the geometric configuration of individual components comprising a distributed physical system."[64]

In this case, the probability omega is "the ratio between the number of possible (spatial) arrangements of system components that can give the current configuration and the total number of possible arrangements yielding all possible configurations of the system."

Deriving an Entropy Measure of Position

Now that position and entropy have been motivated, this section will present the derivation of the entropy measure of position.

It should be clear that the definitions for thermodynamic and configuration entropy only work for more states (i.e., options). What about better options?

For example, configuration entropy only measures access, not desirability, but the thermodynamic entropy incorporates the idea of energy and its dispersal throughout the system of potential microstates of the substance. This brings up an interesting idea. If the energy measure for desirability is incorporated into the expression for entropy, then there is a requirement that the better the options, the higher the entropy might be ensured. The challenge is to find a way to accomplish this goal.

Recall Gibbs entropy, Equation 3.13 has been shown to conform to the energy analogy key to this formulation.

$$S = -k_B \sum p_i ln(p_i) \tag{3.13}$$

Boltzmann's constant is troubling, as it is clearly specific to the realm of thermodynamics. The constant kB relates the energy at the micro level with the temperature at the macro level. For the purposes of this thesis, with no information to indicate otherwise, let $k_B = 1$. The application of this assumption produces Equation 3.14 that looks just like information entropy:[65]

$$H(X) = H_n(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i ln(p_i) \tag{3.14}$$

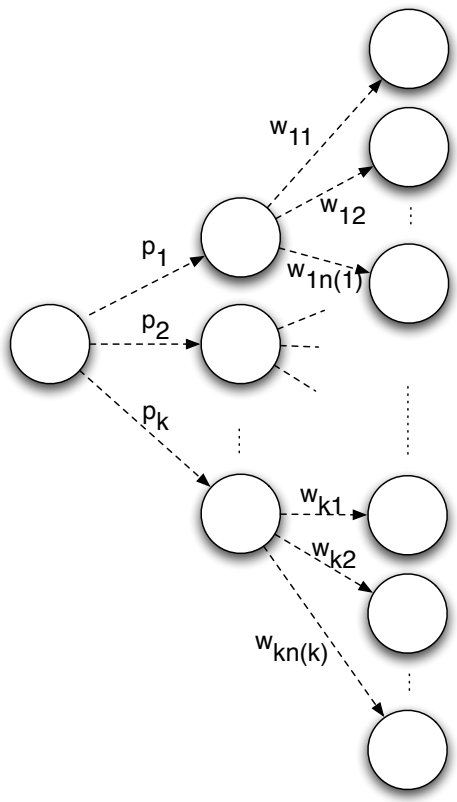


Figure 3.10: A generic system of individual boxes used to derive the strongly additive property of information entropy. The property states that the entropy of a whole ensemble of boxes should be equal to the sum of the entropy of the system of boxes and the individual entropies of the boxes, each weighted with the probability of being in that particular box.

The fact that the trail leads to information entropy is very interesting. There is a mathematical property of information entropy that has direct application to the distributed, interconnected nature of the architecture transition space. This property is known as the strongly additive property of information entropy.

Strongly Additive Property of Information Entropy The strongly additive property of information entropy is derived by examining a system of individual boxes, as shown in Figure 3.10, and finding an expression for the entropy of the whole ensemble of boxes. The property states that the entropy of a whole ensemble of boxes should be equal to the sum of the entropy of the system of boxes and the individual entropies of the boxes, each weighted with the probability of being in that particular box.[66]

Mathematically: Given an ensemble of elements that are divided into k boxes with n(k) elements,

the entropy of the ensemble should be the following:

$$H(p\omega_i) = \sum_{i=1}^{k} [-p_i log(p_i)] + \sum_{i=1}^{k} p_i \sum_{j=1}^{n(i)} [-\omega_{ij} log(\omega_{ij})] \tag{3.15}$$

where k, n are positive integers; $p = (p_1, p_2, ..., p_k)$, $\omega_i = (\omega_{i1}, \omega_{i2}, ..., \omega_{in(i)})$;

$p_i, \omega_{ij} \geq 0$ for all i,j, s.t. $\sum_{i=1}^{k} p_i = 1$ and $\sum_{j=1}^{n(i)} \omega_{ij} = 1$ for all i,n(i).

The concept of the strongly additive property is appealing, but information entropy measures uncertainty, not desirability. According to information theory, the uncertainty is the information conveyed by revealing that X has taken on the value x_i in a given performance of the experiment.[67]

The uncertainty associated with the event $X = x_i$, i = 1,2,...,n, is $\log(h_i)$. Substituting into the expression for information entropy:

$$H(X) = -\sum_{i=1}^{n} (p_i) log(p_i) = -\sum_{i=1}^{n} (p_i) \log(h_i) \tag{3.16}$$

with probabilities $p_1, p_2, ..., p_n$ respectively,

such that $p_i \geq 0, i = 1, 2, ..., n$; and $\sum_{i=1}^{n} p_i = 1$.

The substitution enables conversion of the equation to the measurement of other features, such as desirability.

Again, by analogy:

Additivity of Long-Term Position Entropy The entropy of the long-term position of a given architecture should be equal to the sum of the entropy of the system of options out of the given architecture and the individual entropies of exercising those options, each weighted with the probability of exercising that particular option.

What does this mean for measuring the position of an architecture?

The entropy of the ensemble of options should represent the desirability of being in a given architecture, as measured by its forward-looking position (think of the chess example at the beginning of

this section). The additivity property looks at the overall structure of the system of options as well as the desirability (entropy) of exercising each of those options.

However, there is still the question of how to relate to our "system of options." How do these terms specifically translate into measuring the number of options and the desirability of those options?

Position Entropy

This section motivates and derives the contributions to the position entropy formulation.

First, it is necessary to identify and motivate the different aspects of the architecture transition space that contribute to a decision maker's view that exercising a change between architectures is desirable. An important aspect of using entropy to measure position is that entropy can be positive or negative, hinging upon whether the input to the logarithm, h_i is greater than or less than unity. The working definition of energy in this thesis assumes that energy less than unity is desirable, and energy greater than unity is undesirable. Thus, it is reasonable to take advantage of the relationship in order to add entropy (desirable component) or subtract entropy (undesirable component). In this sense, entropy acts as a reward/penalty accounting system. The identification and motivation of the different components of position entropy is thus framed in the context of rewards and penalties.

Contributions to Desirability This thesis proposes four contributions to desirability as measured by position entropy. These four components appear to be the minimal set based on the structure shown in Figure 3.11.

One of the big "-illities" of system design floating around in the literature is flexibility, the ability of a system to be easily changed.[24] Flexibility can be defined as the number of options that can be exercised immediately since the more options there are, the greater likelihood the system may be easily changed. In the absence of additional information, the more immediate options there are, the greater the flexibility to adapt to external circumstances, the greater the value of the current architecture from the perspective of flexibility. So, there should be a term in the entropy equation that expresses the reward for the amount of flexibility inherent to the system in question. Since the number of options is a positive integer, there can be no penalty (negative entropy contribution) side of flexibility. Instead, the reward function for flexibility is monotonically increasing with the number of immediately accessible options. It is interesting to note that this definition of flexibility coincides

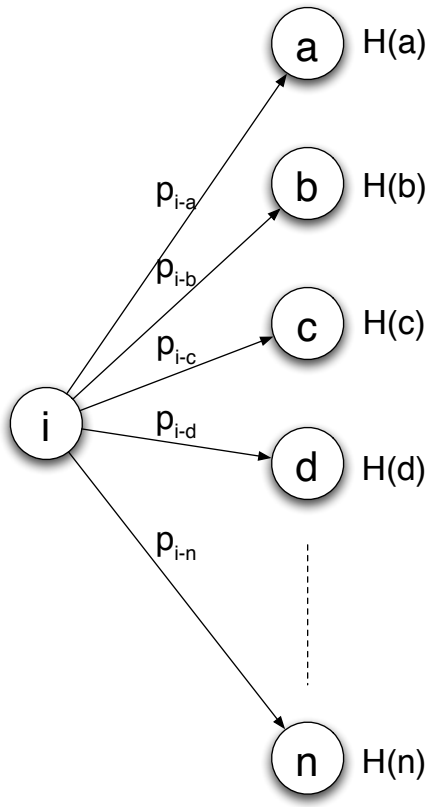


Figure 3.11: Generic position entropy structure. Position entropy is a function of the system of options, the desirability of transitioning from the current state i to a new state (e.g., state a, state b and etc.), the desirability of the new state and the desirability of the future given the new state (e.g., H(a)).

with the definition of configuration entropy, and behaves as a measurement for the desirability (i.e., entropy) of the system of options (i.e., boxes).

If it is desired to impart a penalty to having fewer options than the current system, it is possible to re-define unity such that unity reflects the number of options of the current system. If the new system has few options, it will have a penalty. If it has more options, it will have a reward. For the remainder of this discussion, it will be assumed that everything is calculated relative to the maintain "as is" option. This isn't necessary, and a later section will describe the method to adjust for this comparison after the entropy calculations have been made.

The desirability (i.e., entropy) of exercising each option (i.e., box) is a bit more tricky to articulate. What aspects of exercising the option influence the decision-maker's perspective of desirability?

Suppose it is estimated that transitioning to option A costs $10,000 while transitioning to option B

costs $25,000 and the cost to maintain the system "as is" is expected to cost $5,000. In the absence of any other information, which option is most desirable? The answer is to maintain the system as-is by a factor of 1:2 to option A and 1:5 to option B. Now, suppose that it is known that the cumulative cost of the system is $1,000,000. An additional $5,000 to maintain the system is a mere 0.05 percent, option A is 0.1 percent, and option B is 0.25 percent increase. The choice is still maintain as-is, but now all choices are less than 1 percent of the total cost to date, which makes them seem more equivalent than looking at the absolute numbers. Now, suppose the current total benefits are $2,000,000 and the expectation is that maintaining the system will bring in $10,000, while option A should bring in $25,000 and option B $50,000. On what basis should the decision be made as to which option is most desirable if the only considerations are the outcomes over the course of the transition (single period with history).

The current rate of return on investment is 100 percent over the course of the systems' history. One can argue that a transition only makes sense if the rate of return is increased (assuming there is no threat-driven reason to change). The desirability of the option is measured relative to the legacy maintain as-is option since this is the default choice and it is necessary to compare results along the same period. Obviously, the new rate of return on investment for the maintain case is 100 percent, which happens to be the same as the rate of return for option B. The new rate of return for option A happens to be 0.24 percent more than the default over the period. If the period is one month, then this is equivalent to an INCREASE of 2.88 percent over a year. Based on this information, transitioning to option A is the most desirable choice. When applied to the entropy formulation, an option equivalent to maintaining as-is will be neutral, no reward and no penalty. If the option is better than the default, it receives a reward commensurate with the strength of the benefit. Similarly, if the option is worse, it receives a penalty commensurate with the strength of the difference.

An interesting outcome of this thinking is a new realization of the meaning of energy. Suppose the energy of a transition — viewed as the cost of transition by the benefits gained in the same period, not accounting for history — is greater than unity. This means that the costs exceed the benefits. The greater this number, the longer it will take to pay off the investment given the current rate of benefit accrual. In a sense, this represents the amount of effort required to achieve a return on the investment. This concept is very similar to inertia. Let this effect be known as transition inertia. The meaning of inertia for overall energy depends on the relative scalings of the legacy costs and benefits to the transition costs and benefits. So, this effect can be referred to as legacy inertia, though

the strength of the effect is dependent on scale.

In any case, the immediate effect of exercising the transition has an impact on the desirability of the option.

Once the transition has been made, the desirability of exercising the option largely rests upon the desirability of the new architecture independent of any other information. In some ways, it does not make sense to have this term in addition to the transition entropy. However, the formulation is accounting for two separate periods. There is the time during which the transition is occurring (see assumptions above), and then there is at least one period in which the new system is operating. It is safe to assume this because there is a natural lag following a transition in order for the decision makers to observe the impact of the change before deciding whether another large change is necessary (this does not count repairs/corrections — these would be part of the new maintenance costs for this second period). Thus, it is important to account for the desirability of the new state independent of the transition period.

Finally, there must be the recognition that the desirability of exercising the option requires considering the desirability of the evolution potential out of the new architecture. Thus far, the formulation has only considered the immediate effects of the change. But as the chess example should make abundantly clear, it is critical to look at the future consequences of making the change. The knight can take the queen and check the king and earn large rewards for doing so, but in the next step, the knight will be eliminated by another knight. This option is in contrast to taking the pawn with small immediate rewards, but large rewards for checking the king without immediate threat. Of course, the second choice would be short-sighted since the king will move out of check and the potential for checkmate will have been wasted. More than likely, the knight should remain where it is and reinforcements should be brought in to secure the checkmate, just like in the engineering world where sometimes it makes sense to maintain as-is until the technology readiness of a key component has matured. But if one were to suppose that these pieces had already been in position, the optimal choice would be to take the pawn rather than the queen. The desirability of the future given the new state is simply the position entropy of this new state looking forward.

It is a theoretical note that that there could be other "-illities" expressed in the formulation. However, some of these "-illities" are addressed by considering the change exposure (for example, robustness), or could be incorporated into the evaluation of energy. A new separate term is likely to appear if there is another "-illity" that is critically linked to the structure of the transitions, like flexibility.

It is also important to point out that, like all metrics of this form, one must be careful to ensure that the strengths of the terms are equivalent. One can see the danger by considering weighted sum metrics. One term may take on more weight (without cause) on the basis of the scale. The impact of scaling is somewhat diminished for this metric because of the definition of energy. However, the combination of the energy-based terms centered around unity with the integer-based flexibility term should raise concerns of unfair emphasis. For this reason, there is the caution to use appropriate scaling.

Along a similar line of reasoning, it is recommended that appropriate weightings be considered for each term, chosen on the basis of how the decision maker values the different contributions to his or her view of desirability. Naturally, if there are multiple decision makers (stakeholders), there are methods available for combining the desirabilities to arrive at a reasonable approximation of the situation.

On the basis of the above discussion, the components to position entropy are given as:

- The system of options.
- Each option if exercised:
  - Desirability of transitioning from current state to new state.
  - Desirability of new state.
  - Desirability of future given new state.

These attributes are explored in Figure 3.11 on page 151 and in the following sections, which mathematically derive the substitutions of h_i for each contribution to the desirability. An estimation for the probability of exercising the option is proposed as well. For the moment, it is assumed that the scaling and weightings are all equal to unity.

Probability of Exercising an Option Estimating the probability that a decision maker will choose a particular option is no easy task. There are a lot of factors that go into making a decision of this kind. One method would be to interview the decision maker to identify preference curves and derive a measure of the probability based on these.

A simpler method, though perhaps not as accurate, assumes an equivalence between the probability that a decision maker will choose an option and the probability that the option is desirable relative to the other accessible options.

If there is no information about the desirability of an option, the probability of exercising the option can be assumed to be:

$$p_{i \to j} = \frac{1}{C(k)} \tag{3.17}$$

where C(k) is the cardinality (number of elements) of the set of options k.

However, there is limited information about the desirability of the options, thus the probability of exercising the option can be refined. Let it be assumed that the greater the value of an architecture j relative to the values of all immediate options k, the greater the probability the decision maker will chose architecture j over the other options. Mathematically, this idea can be expressed as:

$$p_{i \to j} = \frac{V_j}{\sum_{j \in k} V_j} \tag{3.18}$$

where $V_j = \frac{B_j}{C_j} = \frac{1}{E_j}$

Clearly, there could be numerous ways to estimate the probability of transitioning, with varying degrees of accuracy. The idea is to model the likelihood that a decision maker will make a particular choice. There are many factors that influence decisions, including politics. There is also much uncertainty as to how the decision maker may view things in the future. The idea is simply to get as close as possible to the likely distribution function for the decision probabilities. In a manner, this thesis contributes a model for the uncertainty inherent in the decision making itself, though the model is by no means the end-all.

Configuration Entropy: A Measure of Flexibility The first term in the position entropy formulation is the expression for the entropy of the system of options, also known as configuration entropy. The configuration entropy measures the entropy of the system of options based on the geometric configuration, which provides a good model for flexibility as discussed earlier.

Recall that configuration entropy is the "ratio between the number of possible (spatial) arrangements of system components that can give the current configuration and the total number of possible arrangements yielding all possible configurations of the system."

The configuration entropy of the current architecture i can be written as:

$$-\sum_{j\in k} p_{i\to j} log(\Omega_j) \tag{3.19}$$

where Ω_j is the ratio between the number of immediate options giving the architecture j and the total number of immediate options yielding all possible architectures.

It should be clear that in this case that:

$$\Omega_j = \frac{1}{C(k)} \tag{3.20}$$

where k is the set of immediate transition options out of the current architecture and C(k) is the cardinality (number of elements) of that set.

Thus, the entropy of the system of options, an expression for the flexibility of the system, is given by:

$$-\sum_{j\in k} p_{i\to j} log\left(\frac{1}{C(k)}\right) \tag{3.21}$$

The cardinality of the set k is equivalent to the outdegree of the architecture as proposed by Ross and Hastings. Recalling Ross and Hastings' methodology brings up an interesting point about acceptable costs of transitioning. This thesis does not assume there is any filtering on this basis, but it is trivial to implement and should be implemented in systems where there is a definite limitation to the amount of capital that can be outlayed during in a period.

State Transition Entropy Here, the expression for the state transition entropy component is derived. First, start with the generic formulation incorporating the h_i substitution, which is used to translate information entropy to a more general form (Note: here, h_i includes the logarithm. The reason for this will become apparent):

$$-\sum_{j\in k} p_{i\to j} h_{i\to j} \tag{3.22}$$

Now, redefine h such that it reflects the desirability associated with transitioning between the current architecture i and a new architecture j in the set of options k.

If the following condition holds, then the entropy contribution from transitioning should be positive (i.e., transitioning is desirable):

$$\left(\frac{V_j}{V_i}\right) > 1 \tag{3.23}$$

Apply a logarithm:

$$log\left(\frac{V_j}{V_i}\right) > 0$$
$$log\left(\frac{\frac{1}{E_j}}{\frac{1}{E_i}}\right) = log\left(\frac{E_i}{E_j}\right) > 0$$

$$\Rightarrow -h = log\left(\frac{E_i}{E_j}\right) \tag{3.24}$$

Substituting this value for h_i into the expression gives:

$$\sum_{j\in k} p_{i\to j} log\left(\frac{E_i}{E_j}\right) \tag{3.25}$$

This expression is equivalent to:

$$\sum_{j\in k} p_{i\to j} log\left(\frac{V_j}{V_i}\right) \tag{3.26}$$

Which can also be thought of as:

$$\sum_{j\in k} p_{i\to j} log\left(\frac{C_i B_j}{B_i C_j}\right) \tag{3.27}$$

This formulation is equivalent to the transition desirability thought experiment outlined in the Contributions to Desirability section.

New State Entropy Here the derivation for the new state entropy is derived. Again, start with the generic formulation with the h_i substitution:

$$-\sum_{j\in k} p_{i\to j} h_j \tag{3.28}$$

Now redefine h such that it reflects the desirability associated with being in the new state j.

If the following condition holds, then the entropy contribution from being in the new architecture should be positive (i.e., the new state is desirable):

$$E_j < 1 \tag{3.29}$$

Apply a logarithm:

$$\log(E_j) < 0$$

$$\Rightarrow h = log(E_j) \tag{3.30}$$

Substituting h_i into the general expression gives:

$$\sum_{j \in k} p_{i \to j} log\left(\frac{1}{E_j}\right) \tag{3.31}$$

Position Entropy Formulation Finally, the formulation for the position entropy of an architecture can be found by summing the four contributions as found above. This position entropy is a measure of the desirability of an architecture.

Let k be the set of immediately accessible architectures (options) from the current architecture.

Let i be the current architecture.

The position entropy of i is composed of the configuration entropy (i.e., measure of flexibility) of the system of options, the entropy of the desirability of transitioning from i to another architecture, the entropy of the desirability of each of the new potential architectures, and the entropy of the desirability of future paths given each option out of i. This formula can be written as:

$$H(i) = -\alpha\gamma\beta_1 \sum_{j \in k} p_{i \to j} \ln\left(\frac{1}{C(k)}\right) + \gamma \sum_{j \in k} p_{i \to j} \left\{\beta_2 \ln\left(\frac{E_i}{E_j}\right) + \beta_3 \ln\left(\frac{1}{E_j}\right) + \beta_4 H(j)\right\} \tag{3.32}$$

where the probability of exercising option j is:

$$p_{i \to j} = \frac{V_j}{\sum_{j \in k} (V_j)} \tag{3.33}$$

such that the value of exercising option j is assumed to be:

$$V_j = \left(\frac{1}{E_j} \right) \tag{3.34}$$

and C(k) is the cardinality (number of elements) of the set of options, k.

α is a debiasing factor to be chosen by the decision makers. Since the configuration term is based on unit fractions, the term will tend to be larger than the other terms in the equation, thus significantly biasing the results of the entropy metric towards larger cardinality options. Thus, a decision-maker defined factor should be used here to ensure the terms are on the same order of magnitude.

γ is a debiasing factor to reduce the influence of the number of option paths, which are artificially scoped when the set of configurations to be considered are reduced.

Each of the β terms is the decision-maker defined weighting for the terms.

Rewriting in terms of the benefits and costs gives:

$$H(i) = -\alpha\gamma\beta_1 \sum_{j \in k} p_{i \to j} \ln \left(\frac{1}{C(k)} \right) + \gamma \sum_{j \in k} p_{i \to j} \left\{ \beta_2 ln \left(\frac{C_i B_j}{B_i C_j} \right) + \beta_3 \ln \left(\frac{B_j}{C_j} \right) + \beta_4 H(j) \right\} \tag{3.35}$$

Again, it is important to note that weightings and scaling functions are all set to 1 for the purposes of this thesis. Future work can determine the effect varying these weightings might have on the outcomes, and perhaps the optimal way of setting these values.

The position entropy definition provides a relative measure for the position of an architecture. It enables a decision maker to identify the preferred transition relative its alternatives on the basis of the desirability of the new option space that emerges.

For the position entropy to be meaningful, three preference conditions must hold. These conditions are expressed below in the form of theorems, the proofs for which are included in Appendix C.

These proofs assume all weightings and scalings are set to unity, so again, future work can evaluate the effect these have on the preference transition points.

Theorem 1 (*magis bene*): If architecture A has *more* transition options than architecture B, with all else equal, then architecture A has the preferred entropy to architecture B.

Theorem 2 (*melior bene*): If architecture A has *better* transition options than architecture B, with all else equal, then architecture A has the preferred entropy to architecture B.

Theorem 3 (*optimus bene*): Suppose there exist 2 paths leading to the same optimal (i.e., lowest energy, positive future entropy) architecture. If the first path achieves the optimal architecture in 1 transition, and the second achieves it in 2 transitions, such that the optimal architecture is the only one with sub-unity energy, and with all else equal, the minimum transition path should contribute greater entropy to the current architecture.

Commentary on the use of Entropy:

Entropy, in the original thermodynamic sense, describes the spontaneous unhindered dispersal of energy away from a concentrated source. Boltzmann extended this definition to relate the structure of the possible microstates of particles in a thermodynamic system to their macroscopic observed heat. According to Fleissner and Hofkirchner,[68] numerous uses for entropy have emerged that diverge from the original intent and connection to the physical world.

The initial divergence appears to have occurred with Shannon's information entropy, which defines entropy without any physical meaning. Rather, entropy is used to measure the information content of a transmitted message. More generally, it measures the uncertainty of a random variable. The maximal uncertainty occurs when the events are equally likely.

Entropy is often used as maximization or minimization functions. For example, the maximum likelihood estimator goal functions in mathematical statistics, and the application of entropy for simulated annealing global minimum searches. Similarly, sociology utilizes entropy to measure social equality while biology quantifies complexity using entropy. But where is the connection to the energy-structural relationship required for pure application of the concepts?

The entropy formulation described in this thesis attempts to link the concept of entropy back to an energy-structure relationship. However, there is still a considerable amount of abstractness. The

energy of a system as described in this thesis does not have a strict physical connection. Rather, it is based on an analogy of the behavior of a system's transition to the activation energy of a chemical reaction. The connection is behavioral not physical. The structural aspect is closer in concept to information entropy than to true physical entropy, particularly since the equation position entropy is based upon is a property of information entropy. Yet, there is a flavor of the original intent of entropy since there is a divergence from considering the equation in terms of random variables and a return to thinking of the entropy in terms of energy.

Still, the entropy formulation does not flow directly from the strongly additive property of information entropy. The spirit is intact but the exact mechanism of dealing with exercising the individual configuration options required some "artistic" license. If there is any issue with the particular preliminary formulation proposed in this thesis — debias and weighting factors withstanding — it will be here. The concept is clearly along the right path since the proofs provided in the appendix demonstrate that the key properties required by the equation are met.

The power of considering entropy as a measure of the position and dynamic desirability of a system configuration will be demonstrated with the discussion of Strategic Advantage Maps later in this section and in the case studies.

A Simple Demonstration

This section will briefly discuss the impact of the horizon time on the entropy calculations, and demonstrate the impact of each the four terms of the equation separately. The network used for this demonstration is shown in Figure 3.12 on page 162. As the effect of varying each term in the entropy formulation is shown, the values for the benefits and costs in the Figure will be updated accordingly.

Assumptions It is assumed that the benefit accrual and cost outlay of each architecture is constant, as is the transition cost for each transition arc. All architectures are retired at the horizon, so the terminal entropy is zero.

Assume that the retirement cost is the same as the operation cost for a given architecture. Otherwise the system is either maintain as-is or it will be transitioned to a new accessible state. The current state continues operating and accruing benefits while the system is transitioned. The architecture is

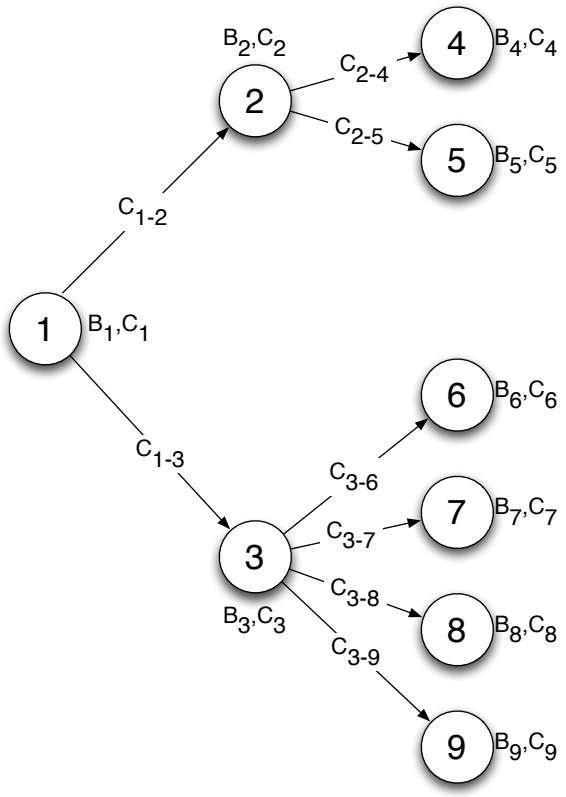


Figure 3.12: A simple demonstration network. The benefits and costs are assumed to remain constant for each time period and are added to the system energy according to the path undertaken and the time period considered.

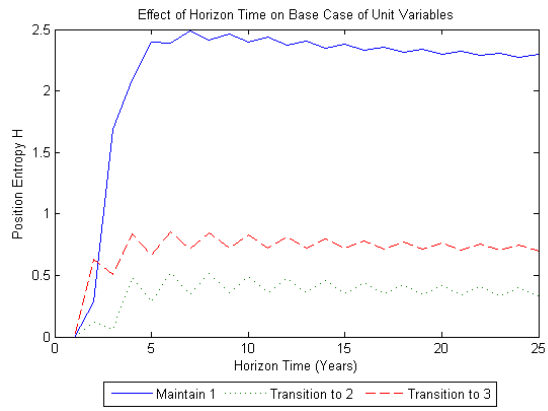
switched to the new architecture at the end of the time period. When the current state is transitioned to a new state, the costs are the transition cost and the operation cost for the current state.

Let it be assumed that all of the debiasing factors are equal to one.

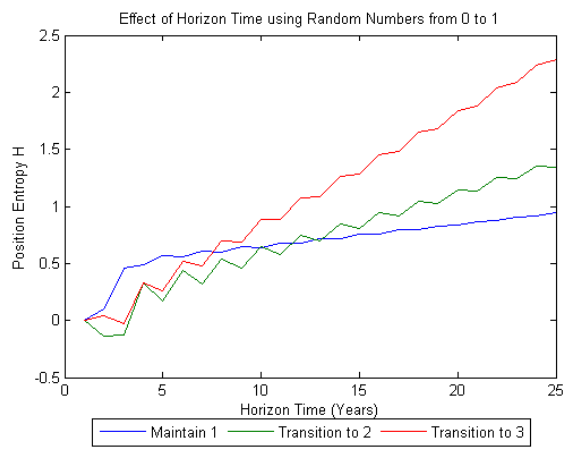
Impact of horizon time First, let's assume that all costs and benefits are equal to one. How do the position entropies change as the horizon time is increased? And let's compare this result to using random numbers from zero to one and see what, if anything, changes.

The base case involving unity variables is shown in Figure 3.13a on page 163. Note that the position entropy values level off and for the most part the relative positions of the options remain constant. For example, the maintain Architecture 1 option is always much more desirable than transitioning to Architecture 3 which is always more desirable than transitioning to Architecture 2.

In the random variable example in Figure 3.13b on page 163, the entropy values increase as the horizon time is increased. Furthermore, the relative positions change. Early on, the maintain Ar-



(a) Unity Variables



(b) Random Variables

Figure 3.13: A plot of ending position entropies over a range of horizon times for the (a) base case with all variables equal to unity and (b) the case in which all variables are randomly chosen on a range of zero to one.

chitecture 1 option is more desirable than either transition option. Then there is a period in which a transition to Architecture 3 is more desirable than the maintain Architecture 1 option, which is still more desirable than transition to Architecture 2. For longer horizon times, the ordering changes again. Now, transitioning to Architecture 3 is more desirable than transitioning to Architecture 2 which is now more desirable than maintaining Architecture 1.

These two examples demonstrate that choosing an appropriate horizon time is important. Choosing a horizon time that is too short may bias unfairly toward maintaining the system as-is.

What is less clear is how to choose an appropriate horizon time. What factors impact the switches in position entropy results?

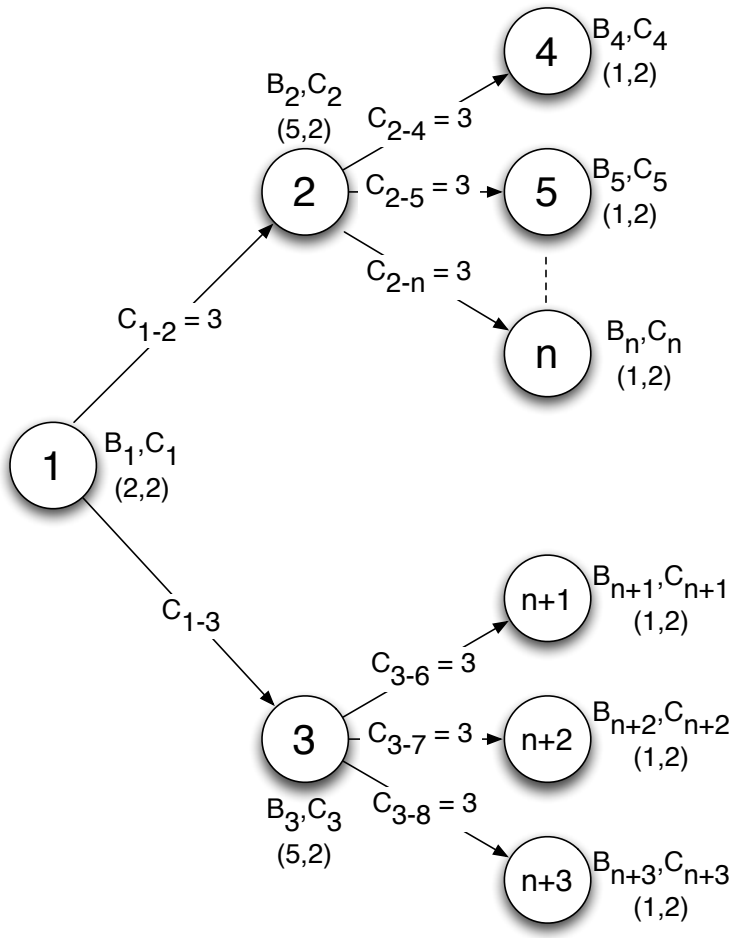
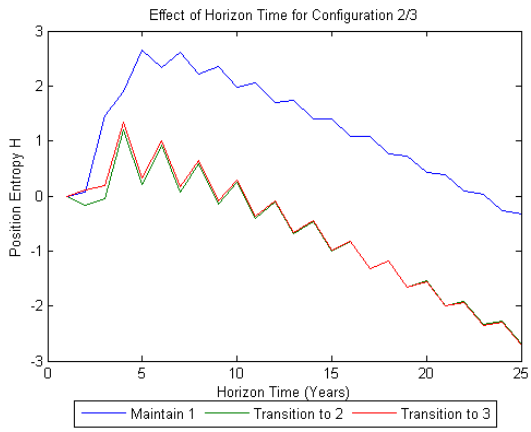


Figure 3.14: The demonstration network for varying the configuration term. Architecture 2 has n outgoing transition options. Architecture 3 has three transition options. The variable n is varied from 2 to 5 to demonstrate the effect of the number of transition options out of Architecture 2 relative to Architecture 3.

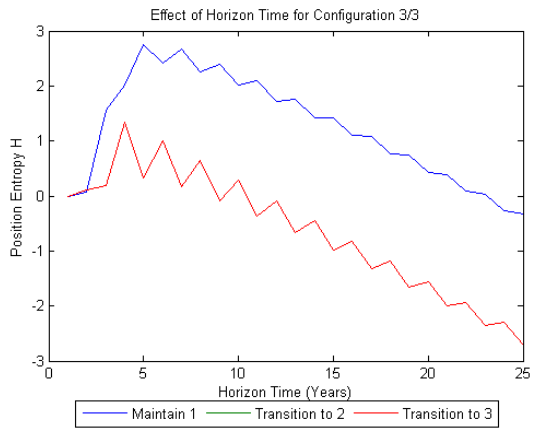
Impact of configuration term Here the impact of the configuration term of the equation is demonstrated.

For this example, the transition costs as well as the benefits and costs of the architectures are held fixed and only the number of options out of Architecture 2 are varied. Architecture 3 is assumed to have three transition options. See Figure 3.14 on page 164 for the values of the benefits and costs.

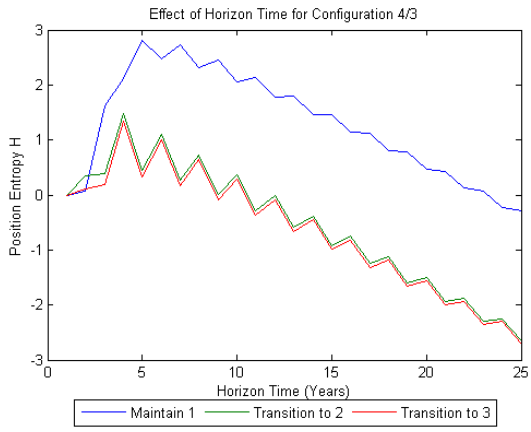
As can be seen by looking at the progression in Figure 3.15 on page 165, as the number of options out of Architecture 2 increases relative to the three outgoing from Architecture 3, the more desirable Architecture 2 becomes relative to Architecture 3. Since all else is equal in the example, the impact of the configuration term can be visualized.



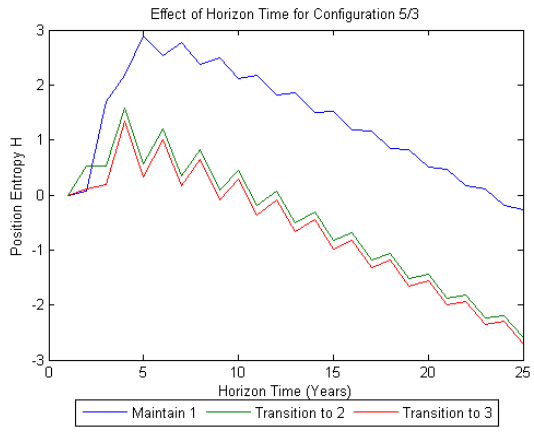
(a) 2/3 Configuration



(b) 3/3 Configuration



(c) 4/3 Configuration



(d) 5/3 Configuration

Figure 3.15: Examining the effect of changing the configuration by varying the number of options for the two transition architectures (Architecture 2 and Architecture 3) and plotting the ending position entropies for all three options for (a) the 2/3 configuration, (b) the 3/3 configuration, (c) the 4/3 configuration, and (d) the 5/3 configuration over a range of horizon times. An x/y configuration refers to Architecture 2 having x outgoing options and Architecture 3 havin y outgoing options.

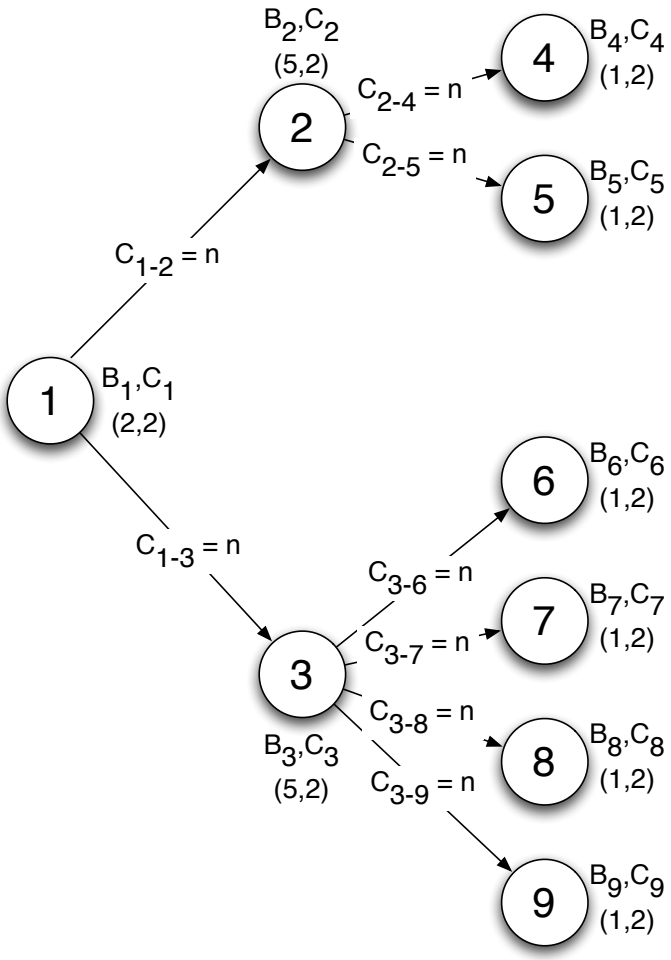
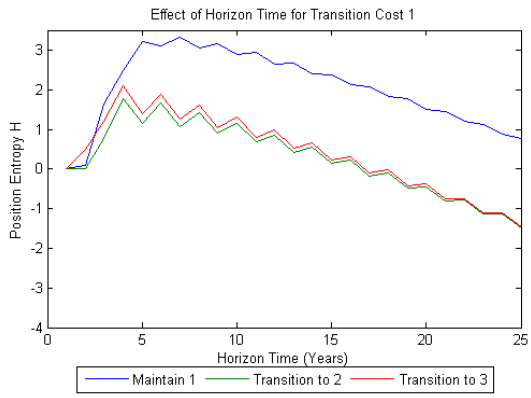


Figure 3.16: The demonstration network for varying the transition term. Here the transition costs are variable, ranging from 1 to 4. The other variables are identical to the demonstration for the configuration term and are held constant for the duration of the demonstration.

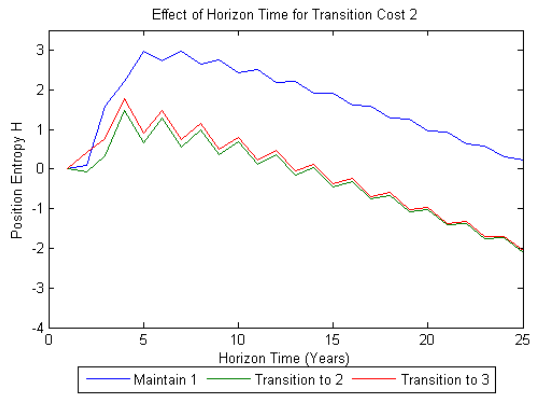
Impact of transition term Here, the impact of the transition term is demonstrated.

Now the number of options and the benefits and operating costs of each architecture are held constant. Only the transition costs are varied. See Figure 3.16 on page 166 for the benefits and costs.

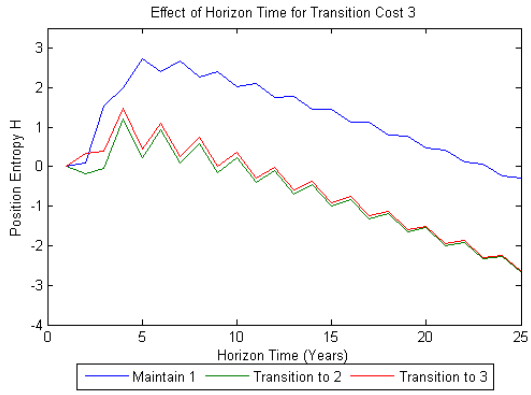
The figures are all to the same scale. The effect of increasing the transition costs relative to the operating costs appears to shift the ending position entropy downwards.



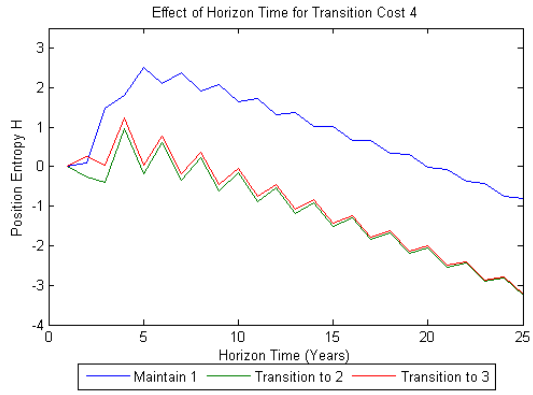
(a) Transition Cost 1



(b) Transition Cost 2



(c) Transition Cost 3



(d) Transition Cost 4

Figure 3.17: Examining the effect of changing the transitioning by varying the transition costs and plotting the ending position entropies for all three options for (a) transition cost 1, (b) transition cost 2, (c) transition cost 3, and (d) transition cost 4 over a range of horizon times.

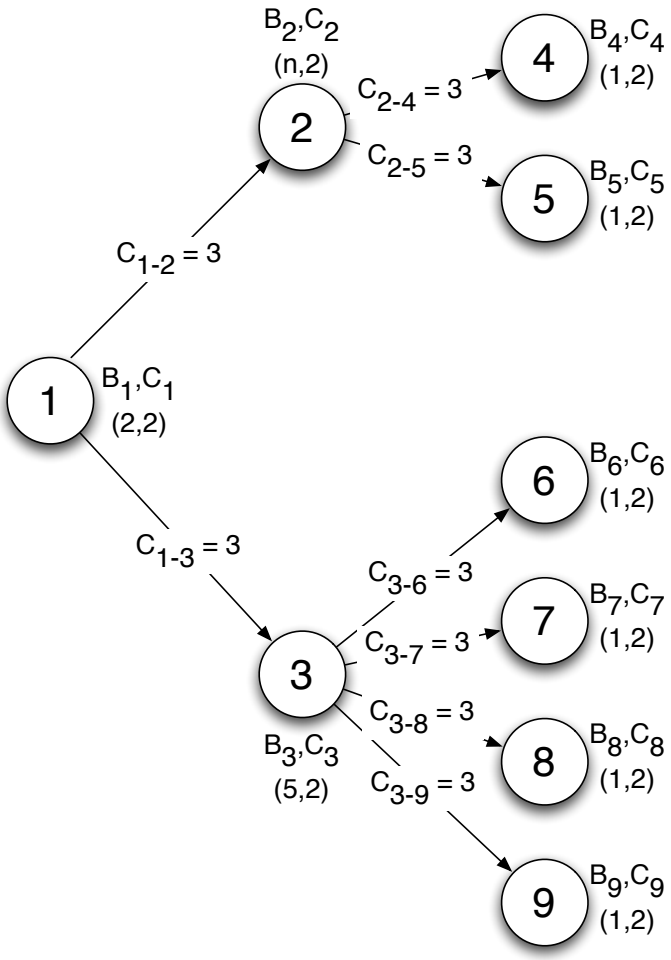
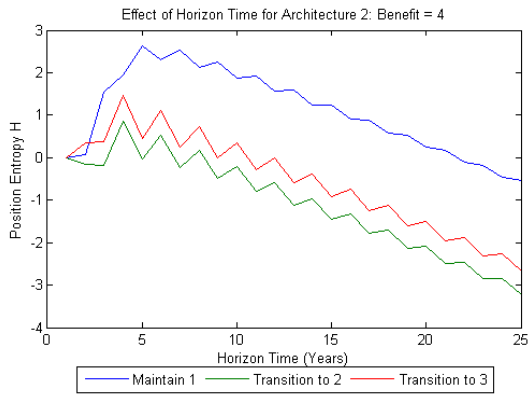


Figure 3.18: The demonstration network for varying the new state term. Here the benefits for architecture 2 are varied, ranging from 4 to 7, relative to 5 as the benefit for architecture 3. The other variables are identical to the demonstration for the configuration and transition terms and are held constant for the duration of the demonstration.

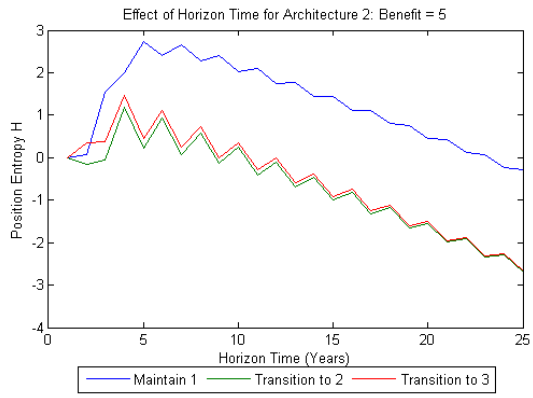
Impact of new state term Here, the impact of the new state term is demonstrated.

To demonstrate the new state term, all variables other than the new state (Architecture 2 and/or Architecture 3) are held constant. See Figure 3.18 on page 168 for benefits and costs.

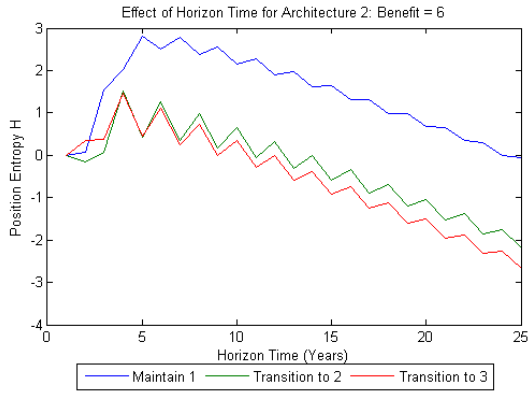
As is evident from Figure 3.19 on page 169, as the benefit of Architecture 2 (i.e., the energy of Architecture 2 decreases and it becomes more desirable) increases relative to Architecture 3, the more desirable Architecture 2 becomes. The effect is enhanced with a longer time horizon. This result makes sense because any difference in benefits accrued in a given time horizon will naturally magnify over time.



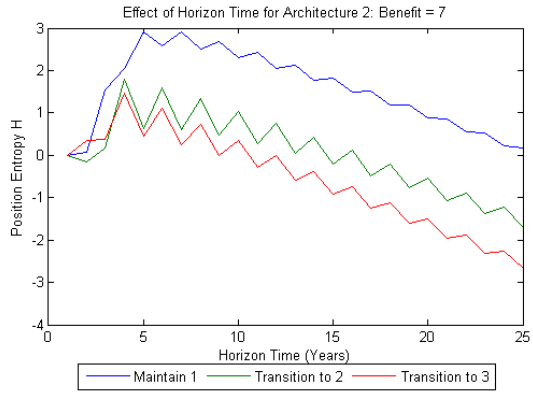
(a) Architecture 2: Benefit 4



(b) Architecture 2: Benefit 5



(c) Architecture 2: Benefit 6



(d) Architecture 2: Benefit 7

Figure 3.19: Examining the effect of changing the new state by varying the benefit of Architecture 2 and plotting the ending position entropies for all three options for (a) benefit 4, (b) benefit 5, (c) benefit 6, and (d) benefit 7 over a range of horizon times.

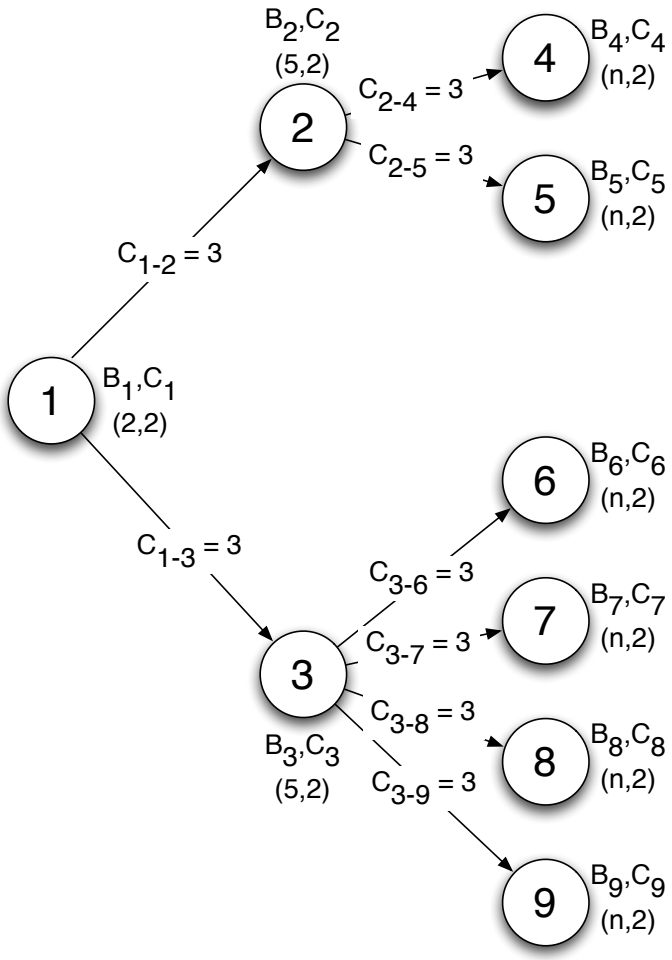
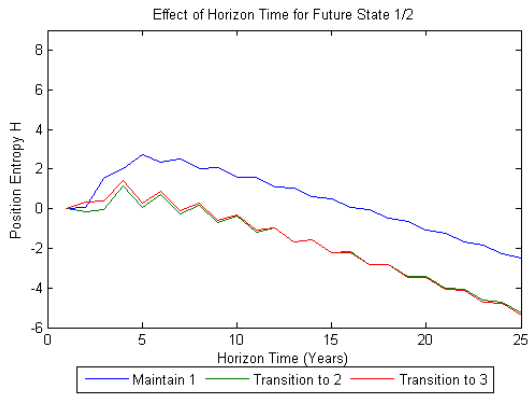


Figure 3.20: The demonstration network for varying the future state term. Here the benefits for Architectures 4 through 9 are varied from 1/2 to 5. The other variables are identical to the demonstration for the configuration and transition terms and are held constant for the duration of the demonstration.

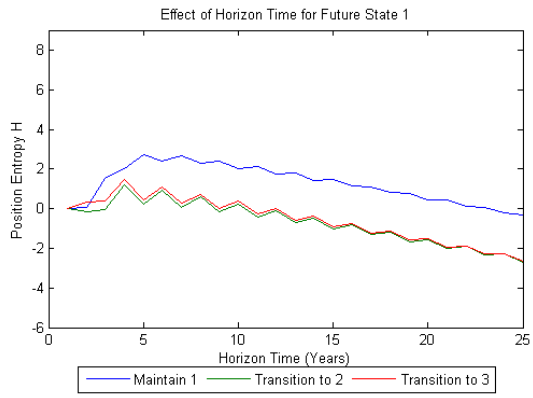
Impact of future term Here, the impact of the future term is demonstrated.

In this case, the goal is to gain insight into the impact of the future terms. To do this, the states accessible from Architecture 2 and Architecture 3 will be varied with all else held constant. See Figure 3.20 on page 170 for the benefits and costs.

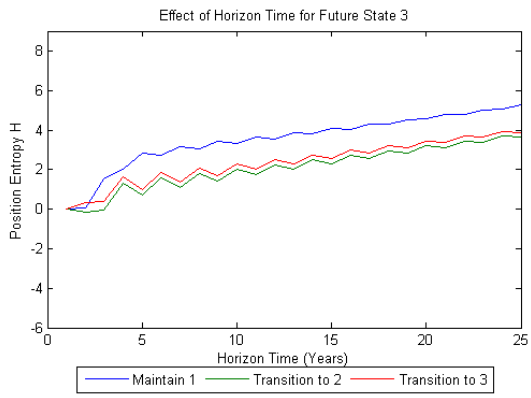
Figure 3.21 on page 171 produces some interesting results. As the benefits of the future states are increased, the "tail" of the graph rotates from dropping entropy with an increase in horizon time to increasing entropy with the increase in horizon time. The drooping tail corresponds with the benefits of the future states being less than the benefit of the current Architecture 1 state. The raised tail corresponds with the benefits of the future states being greater than the benefits of the current



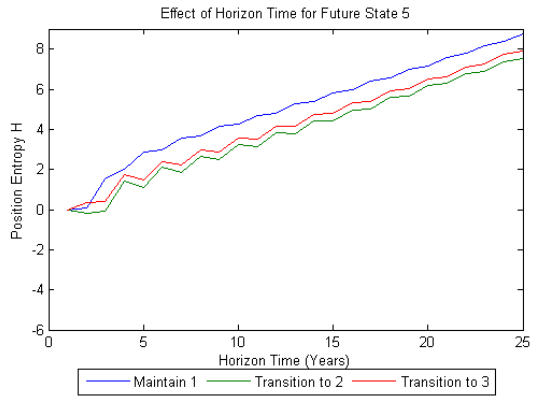
(a) Future State Half



(b) Future State 1



(c) Future State 3



(d) Future State 5

Figure 3.21: Examining the effect of changing the future state by varying the benefit of Architectures 4 through 7 and plotting the ending position entropies for all three options for (a) future state 1, (b) future state 3, (c) future state 5, and (d) future state 7 over a range of horizon times.

state. The transition architectures appear to switch places when the benefits of the future states are less than the current state, but they do not when they are greater.

The other item of note is that the entropy difference between maintaining the current Architecture 1 and the transition architectures 2 or 3 decreases with increasing future benefits.

3.4.4 Transition Entropy

This section describes the meaning of transition entropy and uses the transition entropy to mathematically define option lock-out.

The position entropy of the current architecture may be viewed as a measure of the value of the legacy system as measured by its current legacy inertia and its future potential within the emerging option space. But how does one determine the best option to exercise?

As noted earlier, each term in the position entropy equation seeks to determine the associated reward or penalty for the expected performance of the system over the appropriate period. When the position entropy of an option at a given time is measured relative to the current system, then the reward or penalty is measured relative to the legacy system. Thus, a positive number would indicate a reward (i.e. opportunity) for making the transition whereas a negative number would indicate a penalty (i.e., loss). A value of zero would indicate that there is no measurable difference between the option and the maintain "as is" case.

(Notes: Maintain "as is" is treated as an option in the implementation of the equation; also, must use the maintain option at time t (current state) to compare to an option to exercise between time t-1 and time t. The reason is that the maintain option at time t-1 does not account for the transition (i.e., operating expenses) over the period and thus does not produce a valid comparison.)

So, let us define transition entropy as the difference in entropies between the current and new states. In other words, it is the change in position entropy when an option is exercised, and a measure of the projected reward or penalty associated with exercising the option. Mathematically,

$$H_T(i \rightarrow j,t) = H(j,t) - H(i,t) \tag{3.36}$$

Option Lock-out

What information does the transition entropy provide? Since the position entropy is a measure of the desirability of the system looking into the future, then if the transition entropy is negative (the current state is more desirable than the new state), then there must be something desirable that is being locked-out. There is a net penalty for exercising the option. Option Lock-out comes in varying

degrees; the greater the value of the lock-out as measured by the transition entropy, the greater the penalty associated with exercising the option.

If the transition entropy is positive (the current state is less desirable than the new state), then there must be an opportunity, a net reward for exercising the option.

By avoiding options that produce lock-out, the decision makers can avoid moving into regions of the option space that exhibit high change resistance, or may even result in there being no feasible transition options available. In other words, the higher the lock-out penalty, the greater danger there is of getting stuck where the only recourse is to retire the system or rip out and build new.

3.4.5 Change Exposures

This thesis proposes the concept of a change exposure, which is similar to the idea of risk exposure used in risk management. The risk exposure is an element of risk management that is defined as the probability of a negative event occurring multiplied by the expected loss if the negative event occurs. This concept can be extended to incorporate the opportunity that can be gained as well as to measure the exposure that a given change induces.

A change exposure is a visualization tool that is meant to map the exposure to uncertainty that making a change to a physical system introduces. These visualization tools allow decision makers to quickly evaluate the impact (risks/opportunities) of change, based on their concept of desirability. This thesis introduces two types of change exposures: the strategic advantage map and the "Iceberg Graph" exposure.

Strategic Advantage Map

The Strategic Advantage Map maps the regions of near-term and long-term positions that exercising each transition creates. Here, an architecture's position can be thought of as the region of entropy space it occupies (evaluated over time and uncertainty). The more dominating this region of position entropy is, the more desirable the architecture. Since the dynamic desirability is affected by the system response to various conditions, it is necessary to run a monte-carlo simulation over the uncertainty. Each possible outcome of the uncertainty is plotted as a point on the map for each possible path through the option network.

Advantage measures the desirability of the position of an architecture relative to current or future accessible architectures.

There are two components of position relative to maintaining the legacy system "as is": the near-term impact of a change on its position, and the long-term impact of a change on its position.

It is recommended to evaluate the Advantage Maps out to a minimum of two transition periods, though preferably three or four transitions to gain information on how the desirability of the change is likely to play out over time.

Evaluating Near-term Position The near-term position of the system estimates the uncertainties over the valid forecast horizon (typically out to a maximum of three years for financial forecasts) and evaluates the initial desirability of the system exposure to a transition relative to maintaining the system "as is". This measure indicates the extent of the "immediate" sacrifice or gain by exercising the option.

One method to evaluate the near-term position is to take the energy extrapolations and evaluate the transition and new state entropies assuming that the intermediate states do not exist. It is not necessary to include the configuration entropy because the structure of the system of options is irrelevant.

For example, if evaluating the near-term position over one transition period such as in Figure 3.22, the probability, transition entropy and new state entropy can be estimated using E_{01} and E_{11} relative to E_0. If evaluating over two transition periods, then the probability, transition and new state entropies can be found using E_{011}, E_{012}, E_{111} and E_{112} relative to E_0, ignoring E_{01} and E_{11} since the effect of these are found in the second transition state.

The idea is to evaluate the effect of the change over some period of time relative to maintaining the system "as is" out to that time period *relative to the current energy*. This allows each path to be treated as a single transition option and enables viewing of the effect of the transition option beyond the immediate time period.

This viewpoint is important because the energy accounting assumes the benefits of transitioning do not kick in until the following time period (since the old system is operating while the transition is occurring and then the new system takes over). Thus, looking at the first transition period charts the effect of the transition cost given the current (legacy) benefits, while looking at the second

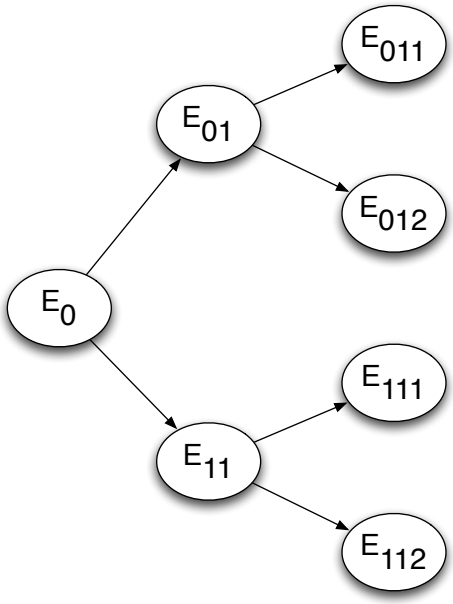


Figure 3.22: Generic system of boxes for evaluating near-term energy.

transition period enables charting of the immediate effect of the new benefits. Looking at only the first transition period obscures the real near-term effect of transitioning. This effect is clear in the example Strategic Advantage Maps shown in Figure 3.23.

Incorporating Long-term Position The long-term position is based simply on the entropy metric for the given architecture at the time period of interest. This value captures the desirability of the architecture looking forward to the terminal values at the horizon. The position is found by evaluating the transition entropy between the given architecture and the maintain "as is" legacy system at the time of interest.

Recall that the long-term position measures the structure of future possibilities weighted by their desirabilities. This measurement indicates the long-term strategic position. Evaluating the difference between this measurement for a change option and the measurement for the maintain "as is" option provides an indication of the degree of penalty or reward for making a change by looking at structural and desirability position from that new architecture.

In order to estimate both types of position, it is very important to measure the entropies relative to the legacy architecture (maintain "as is" option) in the same time period. No matter what path is taken, the energy is always changing. If the entropies are not measured relative to the legacy in the same time period, then the measurement ignores the effect of maintaining the system over that time

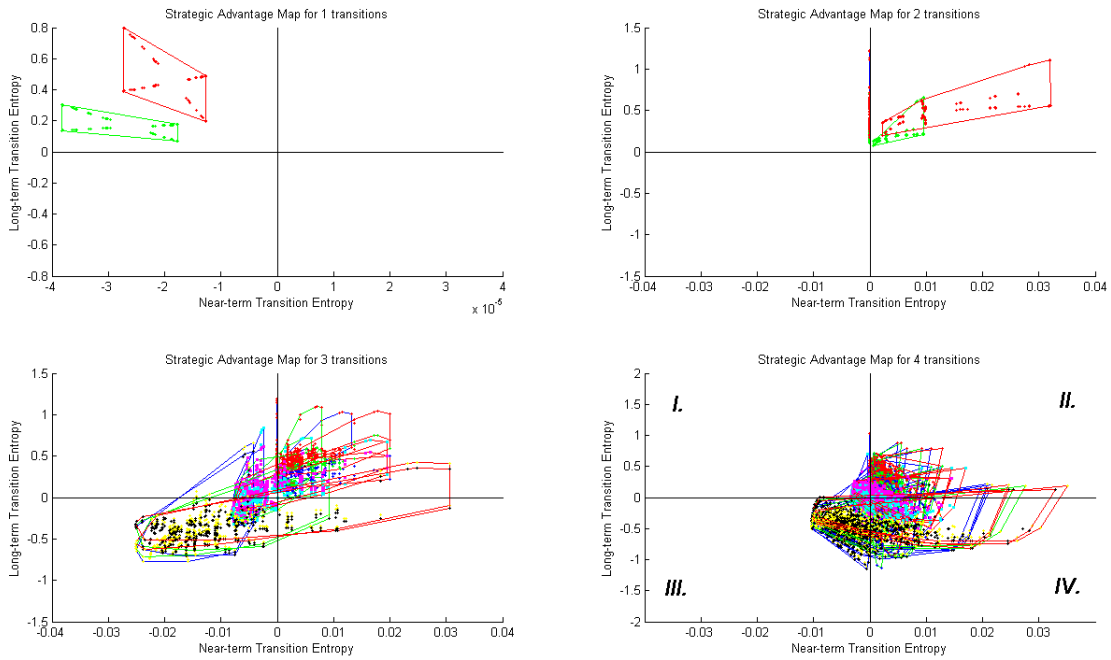


Figure 3.23: Example Strategic Advantage Maps. The map quadrants explained in the text are labelled in the bottom right graph.

period and the relative position is no longer accurate.

Strategic Advantage Map The strategic advantage simply maps the near-term and long-term position measures onto an exposure space in the form of a Hilbert space.[69] If the measures are aggregated over the objectives and requirements, then the exposure space is merely the x-y plane. The origin is always the value of the legacy maintain "as is" system.

Since Monte Carlo is used to evaluate the effect of uncertainty, the near-term and long-term position coordinates map out regions over the advantage space.

Each quadrant of the planar Strategic Advantage map has significance:

Quadrant I: *Sacrifice short-term gain for the long-term advantage.*

Sacrificing the short-term gain for the long-term advantage is equivalent to sacrificing a knight in order to eliminate the opposing queen. A real-life example might be JPL operational strategy for the Pioneer program. In order to ensure the Pioneer ground network would be able to grow into a more long-term network, JPL had to make several sacrifices in the near-term. For example, the desired

antennas could not all be manufactured in time. Only one antenna would be built and operational by the time the JPL Pioneer probes were launched and diplomatic hurdles were making location siting difficult. To compensate, JPL adjusted the operational strategy. The operations schedule was redesigned such that the lunar arrival of the probe would coincide with the lone antenna's line-of-sight. Without an earth-based transmitter, JPL decided to reduce the operational complexity by simply flying by the moon to take photographs rather than attempt a lunar orbit insertion. The plan for the Pioneer probe was non-optimal but it enabled the team to build up the envisioned evolvable network. Please refer to Section 2.1.4 for a more detailed discussion.

Quadrant II: *Short-term gain and the long-term advantage.*

Consider the initial choice of antenna design in the Deep Space Network as an example of achieving the short-term gain and the long-term advantage. The initial Deep Space Network antenna design was chosen to meet the needs of the Pioneer program while also maximizing the ability of the system to grow and expand to meet future needs. As an example of design features meant to enable growth potential, JPL leveraged the opportunity to increase the operating frequency for the Pioneer tracking network from the 108-MHz the Vanguard and Explorer satellites used to 960-MHz. The engineers at JPL recognized that the growth potential of their network would be significantly limited below 500-MHz due to radio noise from terrestrial and galactic sources. Please refer to Section 2.1.4 for a more detailed discussion.

Quadrant III: *Neither short-term gain nor long-term advantage.*

A real-world example of achieving neither the short-term gain nor the long-term advantage is the choice of Vanguard over Orbiter. Placing the Vanguard example here is likely controversial. One could reasonably argue that the choice of Vanguard was designed to provide the U.S. with the long-term advantage as a result of establishing the freedom-of-space precedent using non-military equipment. It was also intended to achieve the short-term gain by way of world-wide acclaim for the first satellite launch. However, the assumptions underpinning these intentions were ill-founded. The design of Vanguard was poorly conceived, its implementation poorly managed and the state of the Soviet rocket program was significantly more advanced than the U.S. thought. Vanguard appeared to be a disaster waiting to happen from the start. Had the designs been evaluated fairly it should have been apparent the great risk that was being undertaken in choosing Vanguard. Although there was

a small region of likelihood placing the choice of Vanguard into Quadrant II as the decision makers apparently assumed, a fair examination would have placed the majority of the region in Quadrant III. Please refer to Section 2.1.2 for more information.

Quadrant IV: *Sacrifice long-term advantage for the short-term gain.*

Sacrificing the long-term advantage for the short-term gain is equivalent to knocking out the opposing knight but losing the queen as a consequence. A real-life example might be the Space Technology Laboratory strategy for the Pioneer program. With less than five months to set up a tracking network, the Space Technology Laboratory had to focus exclusively on meeting the needs of the Pioneer program, including choosing station locations strictly for their favorable look angles for transmitting commands to insert probes into lunar orbit. The newly-built Jodrell Bank antenna was chosen to receive data from the probes. When STL and JPL competed for control of the ground support systems to meet the needs of future ARPA programs, STL lost the bid since their legacy decisions seemingly drove the estimated costs to $34 million combined with an unrealistic timeline for completion. Please refer to Section 2.1.4 for more information.

Commentary It is extremely important to understand that Advantage is in the eye of a beholder. How the options are valued relative to each other is a construction of the decision maker, based on experience and the relative importance of the various objectives. The definition of energy is an art form. The energies must be incremented and valued equivalently among the options or the Strategic Advantage Map will not work.

The Advantage Map does not capture the exposure experienced by the objectives as a function of the contributing uncertainties. Also, the Advantage Map is based on the energy ratios and thus does not capture the scaling of the system.

The primary disadvantage of the Map is the effect of bias error. If biases are not correctly balanced, the Map can produce very different answers. Sources of bias error include the inherent biases of the equation (it is important to balance these biases appropriately), and the relative importance of the equation components (decision-makers should be very careful when estimating the weightings). The equation inherently biases toward Source regions of the option space and against Sink regions. If the Sink regions are truly sinks, then this can be a very good thing because the Map automatically tends against the significant risk posed by option Sinks. However, if the Sink architectures are scope

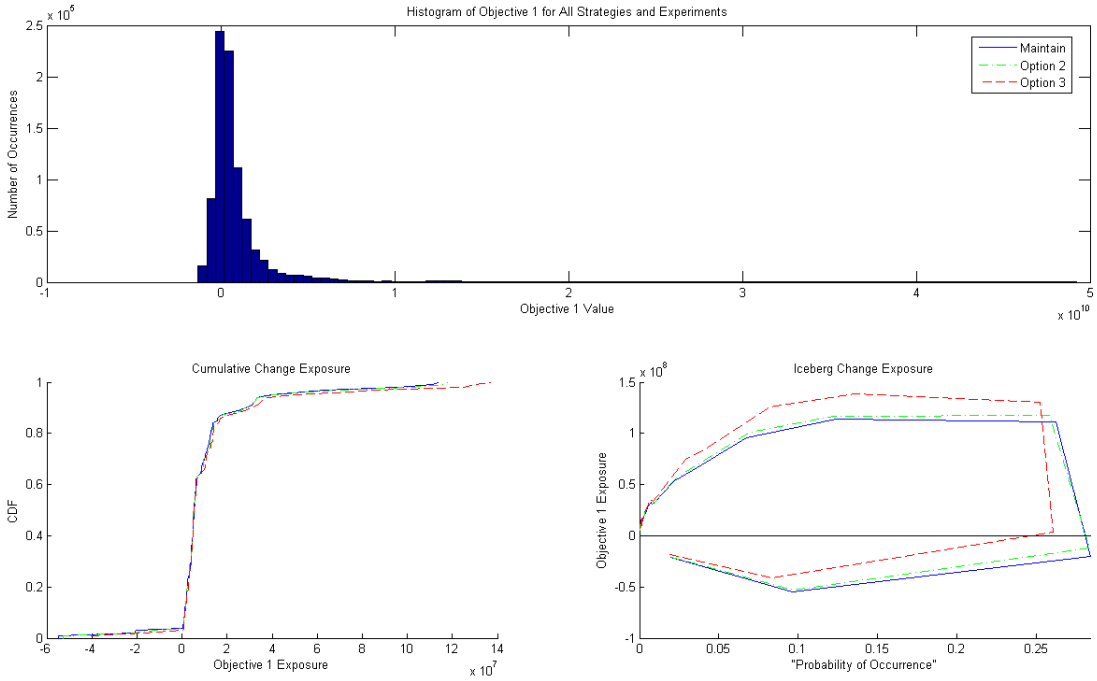


Figure 3.24: Example monetary-based Iceberg Exposure with comparison to the underlying histogram and CDF.

imposed, in order to artificially reduce the option space, this bias can lead toward a non-optimal result.

It is thus strongly recommended that future work include methods to efficiently and effectively debias the equation from the effect of artificially scoping. Real-world applications of this method will need a way to calibrate the Map.

Iceberg Exposure

For monetary-based systems, the change-exposure tool includes an "Iceberg Graph," which maps the exposure of net present value for each accessible transition option relative to a neutral no-gain-no-loss line, resulting in a graph resembling an iceberg.

Here, the concept of a change exposure is introduced that is meant to be used in conjunction with the Strategic Advantage Map. The change exposure provides a measure of the scale of the outcomes, as well as a way to map the objectives to its exposure to uncertainty (i.e., what is being risked by making this transition). For each objective affected by uncertainty, by use of the monte carlo method to estimate the effect of uncertainty, a histogram of the outcomes can be produced such as in Figure

3.24. This histogram can be converted to an approximate Cumulative Distribution Function (CDF) of the exposure for each immediate transition option. In the example exposure graph, there are three immediate options: maintain the system "as is", option 1, and option 2. For each histogram bin, the number of occurrences within that bin can be converted to an approximate probability with the center of the bin representing the value that the objective takes.

In the example, the goal is to maximize the objective and its positive exposure (e.g., the Net Present Value (NPV) of a system). It is thus desired to choose a transition that sits as far to the right on the CDF curve as possible. But it is very difficult in this to get a sense of the nature of the risk and opportunity exposures as the curves sit very close to each other. It would be nice to have a graphic that very quickly describes these relationships.

Let's introduce the concept of an "Iceberg Change Exposure," shown in the bottom right graphic of the example. The figure looks like an iceberg sitting in the water. The region below the black water line represents the risk exposure of the system (similar to the unknown and threatening extent of the iceberg below the water). The region above the water represents the opportunity exposure of the system. Each immediate transition receives its own line so it becomes very easy to see the risk and opportunity exposures of each transition. It is clear from this example that option 3 is the best choice from the standpoint of its exposure.

3.5 Plan

Plan behaves as a roadmap decision-making process. The decision maker should be able to use the position information as viewed from his perspective and develop a strategic course of action. It may be clear from the position information that maintaining the current system for another period in order to wait for some uncertainty to resolve may make the most sense. The position information may also indicate that an immediate transition to another configuration with low-level preparations for a future transition is the most strategic plan. In other words, using position and perspective to build a plan enables the decision maker to put together a series of actions with an indication of the approximate level of resources to expend in the next time period.

The plan aspect of strategy is under the purview of the decision makers, and methods exist for technology roadmapping[28]. There are several implications of the theory of the Strategic Evolution of Systems for strategic planning and decision-making. A discussion of these ramifications is included

below.

3.5.1 The Value of Legacy

The Strategic Advantage Map enables an immediate view of the value of the legacy system. It is difficult to accurately value a physical, non-market-traded system. This thesis provides a way to qualitatively value a physical, non-market traded system by evaluating its desirability-based position relative to its accessible alternatives.

If the accessible architectures all fall within Quadrant II, then the legacy architecture is the least desirable architecture and is thus not highly valued. If the accessible architectures all fall within Quadrant III, then the legacy architecture is the most desirable architecture and is thus very highly valued. Similarly, for Quadrant I, one could say the legacy is highly valued in the near term but not well valued in the long term. For Quadrant IV, the legacy is not well valued in the near term but highly valued in the long term. Naturally, this description depends on the number of transition periods one is looking out. It is important to note the evolution rules and energy accounting rules that are in place that would impact the results. Thus, a true accounting of the value of the legacy architecture would need to look across several time periods to debias the effects of the evolution and accounting rules.

Currently, the value of legacy is qualitative, based on a visual inspection of the Map over several periods. However, future work could look at ways to quantitatively value the legacy based on these maps, perhaps as a weighted average of the transition entropies.

3.5.2 Evaluating when to Rip-out and Build-New or Retire

The Strategic Advantage and Change Exposure Maps provide considerable information that can be used to evaluate when certain decisions should be made. It is largely assumed that the options to retire or rip out and build new are explicitly considered at the horizon (for example, the expected length of time before hardware obsolescence). Thus, the explicit decisions considered only include transition options or maintaining the legacy system. This doesn't have to be the case, and further refinement of the method can incorporate these decision options at each time period.

Even with the explicit consideration of the rip out and build new or retirement options only at the horizon, it is still possible to use the Map to evaluate whether this should be done earlier. If

the steady-state Strategic Advantage Map shows that all of the options fall within Quadrant III and the Change Exposure indicates that all of the options, including maintaining the system, falls completely or largely within the risk side of the exposure, then there is no way the system or any of the accessible options can reliably achieve the objectives. At this point, it is recommended that at least one rip out and build new option be added to the space, and the positions be re-evaluated. This process can be repeated as needed. If the Advantage Map response is the same as before, then it is probably best to retire the system. If a rip out and build new option demonstrates a strategic advantage, then the system should be ripped out and built new ahead of the planned obsolescence.

3.5.3 Option Identification and Creation during System Operation

Options can be created during the operation of a system in several ways:

- Increase the acceptable cost threshold for included transitions to increase the number of accessible options.[58]
- Invest in research and development to generate a new transition rule within the current cost threshold.
- Develop methods to reduce the transition costs.[15]
- Add accessible systems to the option space (i.e., reduce the artificial option scoping, or think outside the box to identify options not considered before, including the use of new technologies and methods).

3.5.4 Identification and Prioritization of Research and Development

The Strategic Evolution of Systems could be used to identify and prioritize research and development projects for a legacy system by creating temporary options based on the Research and Development (R&D) projects. For example, let Project A be a project geared toward developing a technology that could be incorporated into the legacy system. The new system incorporating the project could be added to the option space, with the uncertainty of the outcome of the project, its costs and its benefits included in the exposure evaluation. The relative positions can be re-calculated and the value of the project can be evaluated in the context of the Strategic Advantage and Exposure Maps. Similarly, multiple projects can be added in and the resulting Maps could help identify the most strategic projects and funding could be assigned on that basis.

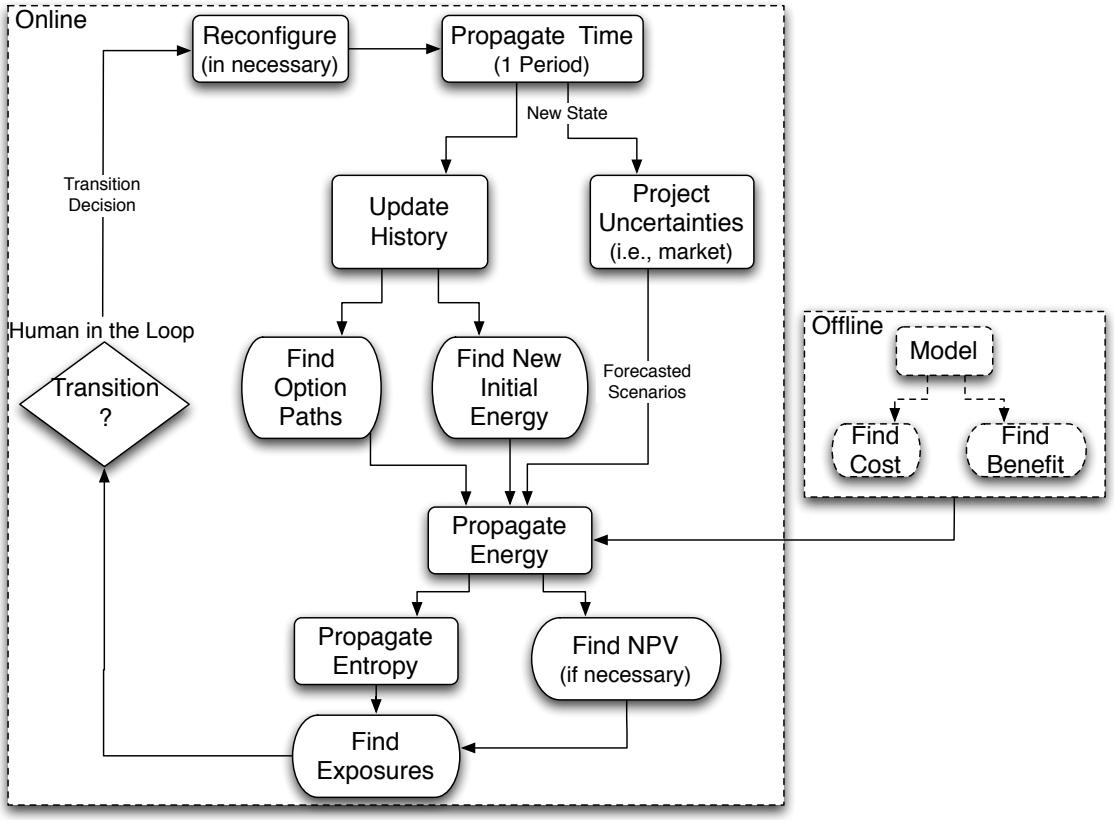


Figure 3.25: Generic simulation model for the Strategic Evolution of Systems framework.

3.6 Pattern

Pattern is the last step of the methodology and reflects the historical documentation and set of decisions along with the impact of those decisions on the performance of the system. As time progresses, this information is updated and used to feedback into the next round of decision making.

Pattern serves to provide continuity, encourage documentation and establish a standard method for evolving the system. These often overlooked aspects of system evolution are especially important for long-life systems.

3.7 The Framework as a Simulation Model

Now that the details of the framework have been described, consider its application to a simulation model. A generic diagram of the Strategic Evolution of Systems simulation model is shown in Figure 3.25.

There are a few important things to note about this model.

First, to implement the framework, there must be a human in the loop to evaluate the change exposures and make a transition decision. It is not possible at this point to automate the process, nor does this author think that full automation is desirable.

Second, time is incorporated by propagating the system one-time period following the transition decision by the decision maker. When time is propagated, the "history" of the system is updated and the new initial energy is calculated based on the "actual" situation over the time period. The transition decision reduces the set of option paths that are now available due to legacy. In order to obtain the system exposures to uncertainty, the outcomes of the uncertainty are forecasted forward to the time horizon (the new starting point of the forecast is the "current" state of the system as determined by the time propagation). This information is used to propagate the energy forward to the horizon over all of the scenarios considered. The energy at the horizon is then used to estimate the terminal entropies, and this information is used as the starting point to propagate the entropies back toward the current time.

Third, the system response is calculated offline prior to running the simulation. Separating the processes can save a lot of time by not having to recalculate the system responses for every possible combination of uncertainty. However, this separation may not be possible for all situations.

This model will be used in the case studies that appear in the following chapters.

Part II

CASE STUDIES

Chapter 4

Global Commercial Satellite System: A Case Study

The following case study is a fictional account based on loosely combining aspects of several existing satellite systems (i.e., Iridium/Globalstar financial difficulties, XM Satellite radio satellite design and services, and etc.) with three transitionable constellation configurations identified and quantified by de Weck et. al.[6] The case study aims to demonstrate the application of the Strategic Evolution of Systems framework for a monetary-based physical non-market traded system.

It is 2009 and there exists a satellite communication corporation that has an inclined-GEO polar constellation providing global mobile radio and data services. The network has been in operation for five years, with an expected satellite lifetime of 15 years. The initial system timeline is shown in Figure 4.2 on page 189. The company policy is to provide customers with comparable quality of service and price to any terrestrial competition.

However, the system is still far from paying off the capital investment of developing, manufacturing and launching the system back in 2004 and investors are getting anxious. The company executives have decided to look into reconfiguring the system in the hopes of capitalizing on a more lucrative market segment. The support staff have negotiated out a number of deals and have lined up a set of potential constellations that have been optimally configured in the traditional way.

Company engineers have constructed a detailed simulation of the system in order to evaluate the likely quality of service the new configurations will provide. The marketing department has compiled competitive monthly user fees for various market segments and has built models to predict

customer growth and usage. The preliminary negotiations have generated an estimate of the likely costs associated with the various configurations.

The executives originally planned to use a standard net present value analysis to determine the appropriate course of action. However, they have recently heard about the development of the Strategic Evolution of Systems framework and decide to run the numbers to see what additional information the method provides.

This case study applies the Strategic Evolution of Systems framework to the problem faced by this company, demonstrates its use in evolving the system and compares its performance over the course of the system lifetime to the standard net present value analysis.

4.1 Perspective

One of the purposes of the perspective portion of the methodology is to ensure that the perspective of the client system is identified, acknowledged and appreciated immediately.

4.1.1 Identify Objectives, Requirements and Constraints

This section describes the objectives, requirements and constraints for the global commercial satellite system.

It should be clear from the scenario in the introduction that the company goal is to achieve a return on the investment. Thus, the primary objective is to maximize the investment return. There are two ways this objective can play out: first, the system turns a profit and the goal is to maximize that profit; or second, the system does not turn a profit and the goal becomes to minimize the loss.

Company policy provides the primary requirement: to provide comparable quality of service and price to the terrestrial competition.

The truly interesting part of this situation is the legacy. For the purposes of simplicity, it shall be assumed that any new satellites launched will be identical to the legacy satellites, although the fuel loads can differ. A problem for future study would be to examine the impact of heterogeneous constellations. This legacy constellation restricts the reconfiguration space. Furthermore, there is the constraint that there are only 10 years remaining for the legacy satellites. There are two ways

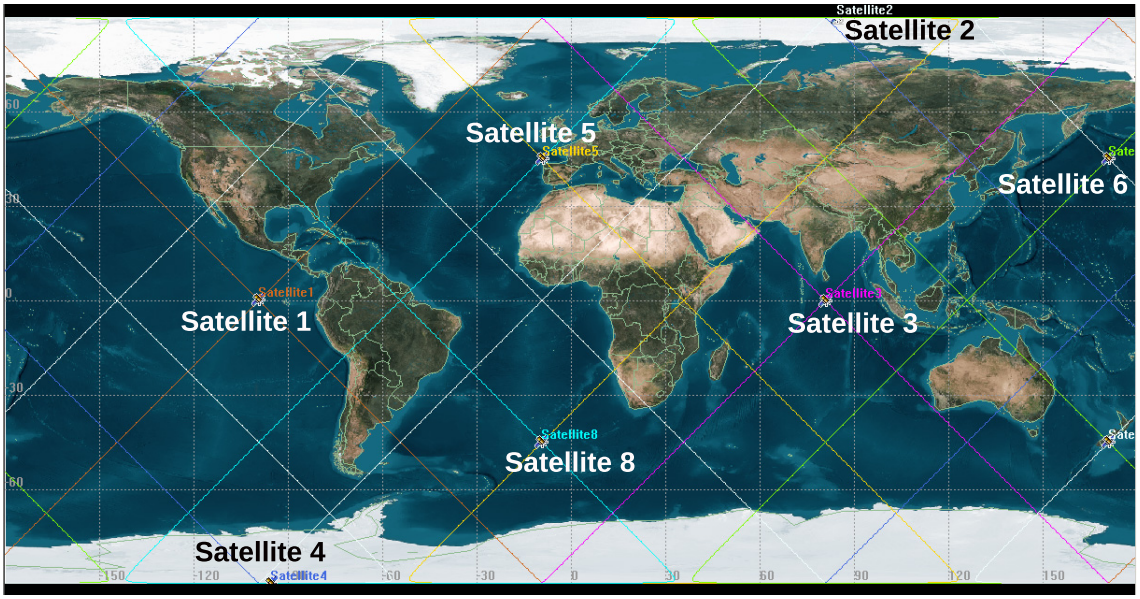


Figure 4.1: Legacy architecture: an inclined-GEO constellation consisted of eight satellites over two planes.

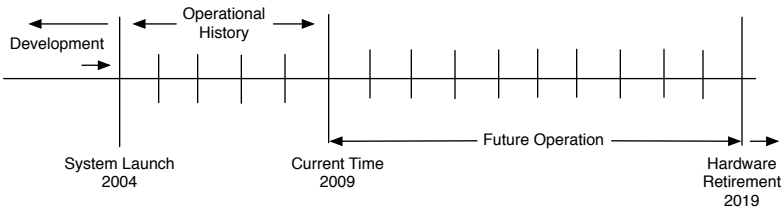


Figure 4.2: Initial system timeline: the current time is 2009. There are five years of operational history and ten years of expected operational time remaining.

to deal with this constraint: pay off the investment within 10 years, or examine what happens when legacy satellites fail and continue operating the system.

Suppose the legacy constellation is a polar, inclined-GEO constellation (35,870 km altitude), consisting of eight satellites over two planes (see Figure 4.1). The satellites have an antenna diameter of 3-meters (ms) and provide 8-Kilowatt (kW) of power for the communications payload. The user device carries the equivalent of a 0.1-meter parabolic antenna. The current network utilizes a form of Interior Gateway Routing Protocol (IGRP) distributed routing, Multiple Frequency Code Division Multiple Access (MF-CDMA) technology, the stop-and-wait ARQ protocol, and Quadrature Phase-Shift Keying (QPSK) modulation. The user data rate is 100 kilobits per second (kbps) with a packet size of 100 bytes.

4.1.2 Identify Decisions, Bounds and Logical Constraints

This section describes the relevant decisions, the bounds on those decisions and the logical constraints between them.

The set of design decisions is provided in Table 4.1. The decisions determine several important aspects of the overall architecture: the design of the constellation, the service parameters accessible to the end user, the size and power of the satellite communications payload and the specifications of the communications architecture that should be transparent to the end user.

The constellation design is specified according to the three logical constraints shown in Table 4.2. It is assumed that the constellations are in inclined-GEO polar orbits. These constellations are plotted along the top horizontal line depicted in the Satellite Constellation Reconfiguration Map, which also identifies the average ΔV required to reconfigure the constellations (see Figure 4.3).[6] The specific constellations considered in this case study are:

- Constellation 1: 8 satellites over 2 planes at minimum elevation angle of 25 degrees.
- Constellation 2: 15 satellites over 3 planes at minimum elevation angle of 35 degrees.
- Constellation 3: 24 satellites over 4 planes at minimum elevation angle of 40 degrees.

The parameters specifying the service characteristics that are accessible to the end-user are the diameter of the receiver on the user device, the average data packet size and the average user data rate. Although service packages do not typically specify the average data packet size, it is implied in the type of service offered (e.g., SMS services use small packet sizes).

The sizing of the communications payload on the satellites is specified by the satellite transmitter diameter and the antenna transmit power.

The aspects of the communications architecture that should be transparent to the end-user include: decisions that specify the degree of centralization of the routing in the system, the modulation scheme and the specific protocols used (i.e., routing, multiple access, and automatic repeat request (ARQ)).

(Note: the InterSatellite Link (ISL) design is an optical 0.2-meter antenna with 100 MHz bandwidth and a total of 4 kW for all channels.)

Morphological Matrix of Satellite Communication Constellation Decisions						
Category	**Design Variable (DV)**	**Decision**	**Units**	**Alt A**	**Alt B**	**Alt C**
Business Strategy	TC	Buy Terrestrial Capacity?	N/A	0 (no)	1 (yes)	
Constellation Design	h	Orbital Altitude	km	10,000	20,000	36,000
Constellation Design	Emin	Min Elevation Angle	Deg	25	35	40
Service	Dr	Receiver Diameter	m	0.1 (Hand-held)		
Service	PS	Average Data Packet Size	Bytes	10	100	1000
Service	Ruser	Average User Data Rate	kbps	10 (Telephony)	100 (Radio/TV)	1000 (Broadband)
Satellite Design	Dt	Satellite Transmit (Tx) Diameter	m	3		
Satellite Design	Pt	Satellite Tx Power	kW	8		
Comm Architecture	ND	Centralized or Distributed Routing?	N/A	0 (Centralized)	1 (Distributed)	
Comm Architecture	RP	Routing Protocol	N/A	1 (Routing Information Protocol (RIP))	2 (IGRP)	3 (Open Shortest Path First (OSPF))
Comm Architecture	MAP	Multiple Access Protocol	N/A	2 (MF-CDMA)		
Comm Architecture	ARQ	ARQ Protocol	N/A	1 (Stop-and-Wait)	2 (Go-Back-N)	3 (Selective Repeat Protocol (SRP))
Comm Architecture	MS	Modulation Scheme	N/A	1 (Binary Phase Shift Keying (BPSK))	2 (QPSK)	

Table 4.1: Morphological matrix for the available global satellite communication system decisions.

Logical Constraints for Satellite Communication Constellation Decision		
Name	**Scope**	**Equation**
Constellation 1	Altitude and Elevation Angle	$((h == 36,000km)\&\&(Emin == 25deg)))$
Constellation 2	Altitude and Elevation Angle	$((h == 20,000km)\&\&(Emin == 35deg)))$
Constellation 3	Altitude and Elevation Angle	$((h == 10,000km)\&\&(Emin == 40deg)))$

Table 4.2: Logical constraints for the global satellite communication system. These decisions define the set of constellations available to the decision makers.

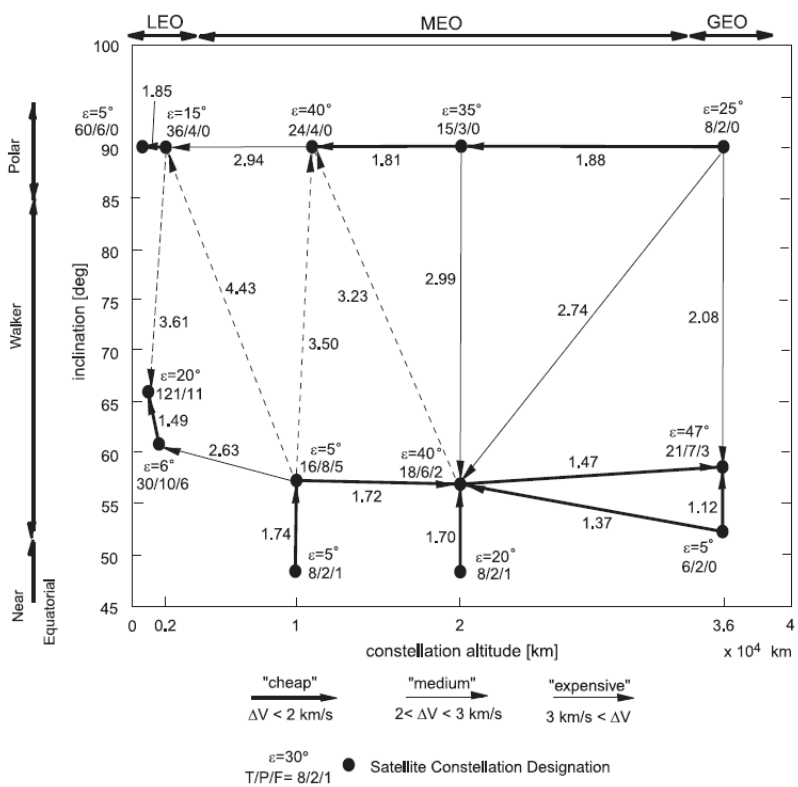


Figure 4.3: Reconfiguration Map, taken from Optimal reconfiguration of satellite constellations with the auction algorithm[6]

Reversible Transition Rules		
DV	**Rule Description**	**Rule**
TC	Adjust terrestrial capacity	↓↑
Emin	Adjust elevation angle	↓↑
RP	Switch routing protocol	↓↑
ARQ	Switch ARQ protocol	↓↑
PS	Adjust packet size	↓↑
Ruser	Adjust data rate	↓↑

Table 4.3: Reversible Transition Rules

Irreversible Transition Rules		
DV	**Rule Description**	**Rule**
h	Change constellation altitude	↓
ND	Switch routing architecture	↓
MS	Switch modulation scheme	↓

Table 4.4: Irreversible Transition Rules

4.1.3 Generate Relevant Architecture Instantiation Networks

Next, the transition rules, supernodes and evolution rules for this system must be determined in order to generate the relevant architecture instantiation network.

Transition Rules

The reversible and irreversible transition rules for this system are shown in Tables 4.3 and 4.4, respectively.

The reversible transition rules involve the terrestrial capacity, the minimum elevation angle, and the protocols and packet size and data rate.

Terrestrial capacity can be leased from terrestrial systems, and can be used to reduce traffic loads on overused links.

Chaize assumes the minimum elevation can only increase in order to guarantee connectivity.[16] However, the simulation model in this case tests area connectivity and will flag constellations that cannot guarantee connectivity.

The protocols and packet size and data rate are all aspects of the system that can be changed via software updates. Therefore, the transitions are reversible but not without cost and risk.

Architecture Family Delimiters			
Decision	**DV**	**Units**	**Value**
Receiver Diameter	Dr	m	0.1 (Hand-held)
Satellite Tx Power	Pt	kW	8
Satellite Tx Diameter	Dt	m	3
Multiple Access Protocol	MAP	N/A	2 (MF-CDMA)

Table 4.5: Architecture Family Delimiters

The irreversible transition rules include the altitude, the routing architecture and the modulation scheme.

So long as the satellites carry sufficient fuel, it is possible to lower the altitudes of the constellation. However, increasing the altitude requires large amounts of fuel, and in the situation considered here, this option is infeasible. The use of orbital tugs might be an option for increasing the altitude. Otherwise, the only other way to increase the altitude is to launch a wholly new constellation. Thus, changing the constellation altitude is considered an irreversible transition.

A distributed routing scheme can be operated as a centralized one by adjusting the routing tables to reflect this kind of architecture. However, it is not feasible to change a centralized architecture to a distributed one as it requires different physical equipment.

Similarly, a Quadrature Phase-Shift Keying (QPSK) system can be run as a Binary Phase Shift Keying (BPSK) system, but not vice versa. Modulation schemes require specific hardware to be constructed.

Those decisions that separate systems into different architectural families are shown in Table 4.5. The diameter and power of the satellites can not be altered once the system has launched, and since the problem assumes a homogeneous constellation, these aspects of the system are fixed unless the system is ripped out and built new.

The user device is considered fixed for simplicity. The antenna size might be varied from device model to model, but this adds a level of complexity that is unnecessary. Future work could look at the effect of heterogeneous user devices.

The multiple access scheme is also assumed to be fixed since each scheme (MF-CDMA or Multiple Frequency Time Division Multiple Access (MF-TDMA)) requires different hardware.

A few more simplifications can reasonably be made:

First, eliminate the BPSK option. QPSK performs better at the expense of needing a higher signal-

to-noise ratio with the user receivers. It may not make sense to invest money in building a QPSK system but running it as BPSK.

Second, eliminate the centralized routing option. Although this option is useful for billing, it adds a substantial amount of unnecessary latency and creates potential congestion issues.

For now, assume that there is no option to lease terrestrial capacity. This aspect of the decision structure does not add to the case study, only adds complexity, but can be looked at in future work.

Supernodes

Now, the only irreversible decision is the constellation altitude. This simplification makes it easy to identify the system supernodes. Figure 4.4 demonstrates the supernodes for the option network of the two highest-altitude constellations.

Supernodes are delimited by irreversible transition rules:

- All architectures within a supernode are fully reversibly connected.
- Supernodes illuminate the underlying relationship structure between the architectures.
- Supernodes can act as sources, sinks or relays.

An initial architecture space reduction should be performed using simulation and filters to eliminate poorly performing designs. In this case study, those architectures that could not meet any of the service requirements were eliminated from further consideration. Employing this step can greatly reduce the computational time required later in the methodology.

Evolution Rules

This case study makes use of five evolution rules. They are as follows:

1. Each architecture is maintained for at least one year before transitioning. This rule translates to each architecture appearing for at least two periods because of the manner in which time is tracked.

2. If it is decided to reconfigure the constellation (reduce the constellation altitude, here), then there must be a lag of three years to account for the manufacturing, launch, and operations preparation that must be made.

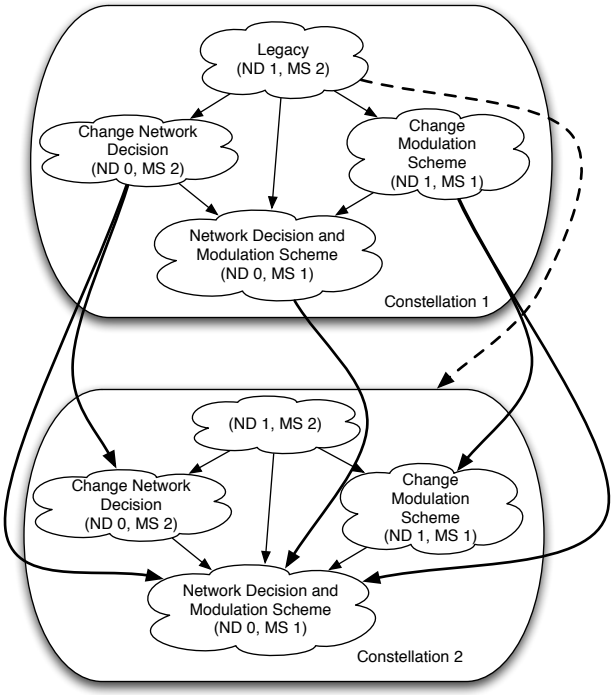


Figure 4.4: Simplified Option Network

3. If only the software is changing, then the transition has the same lag as for maintaining. In other words, the software changes can be written, tested, and implemented within one transition period (year).

4. If the same constellation is maintained as-is until the horizon, then it must be retired at the horizon.

5. If the same constellation is maintained as-is and then transitioned at the horizon, then it is assumed to be ripped-out and the transitioned architecture built new at the horizon time.

4.1.4 Estimating System Desirability

This section describes how the system desirability is estimated.

The case study focuses on a commercial system with the primary objective being to maximize investment return. The greater the projected investment return, the greater the desirability of the system. The smaller the ratio of the cumulative costs to the potential cumulative revenue of the system, the greater the return on the investment.

The current desirability of the legacy system, also known as the initial energy, is the ratio of the cumulative actual costs to develop, manufacture, launch and operate the system of satellites to the present day to the cumulative actual revenues over the operational life of the satellite.

When looking toward the future, it is necessary to consider the potential revenue and any additional costs associated with the further operation and reconfiguration of the constellation. The potential revenue is based on market research and the expected services that the constellation configuration can provide. The additional costs to be estimated include the costs to make transitions between the reversibly-connected states, the costs associated with physically reconfiguring the constellation and launching additional satellites, and the costs to continue operations.

4.1.5 Estimating System Costs

The cost estimations for the events that can occur during system operation include operational costs, irreversible reconfiguration costs and reversible reconfiguration costs. Contingency expenses are not considered here.

Operational costs are essentially the costs required to maintain the system as-is. The operational cost includes the maintenance cost required for each ground-based gateway plus the estimated labor cost to perform the maintenance and to monitor the space segment. These costs are based on SMAD chapter 20 cost estimations. For simplicity, it assumed that each year's operations costs is the operations cost in 2002 adjusted for the current year's dollars.

The irreversible reconfigurations are simply altitude and minimum elevation angle adjustments. These transitions include costs to manufacture the required satellites for increasing the size of the constellation, including the new orbital spares, estimates on the program-level cost associated with making such large-scale changes, the launch operations and orbital support cost, the cost to launch all of the new satellites (assuming at least one launch per orbital plane under adjustment), and the cost to build the additional required gateways. The cost estimations assume that the original satellite design is fully leveraged, so no additional development cost, lower program level costs, or additional software costs are incurred. The fuel expenditures required for lowering the altitude and adjusting the elevation angles have already been estimated in Optimal Reconfiguration paper.

Note: this case study assumes that the gateways are appropriately located, even though the lower constellations place planes in different longitudes. This would be an area for future study.

Reversible Transition Costs			
Level	**Heritage Factor**	**Transitions**	**Total Cost**
1	0.8	Ruser + RP + ARQ + PS	$3,059,700
2	0.7	Ruser + RP + ARQ or RP + ARQ + PS	$2,677,200
3	0.6	Ruser + RP + PS	$2,294,800
4	0.5	RP + ARQ	$1,912,300
5	0.4	Ruser + RP or RP + PS or Ruser + ARQ or ARQ + PS	$1,529,800
6	0.3	Ruser + PS	$1,147,400
7	0.2	RP or ARQ	$764,920
8	0.1	Ruser or PS	$382,460

Table 4.6: The reversibly-connected reconfiguration transitions include the software changes to implement automatic repeat request (ARQ) and routing protocol changes and to adjust the user data rate and packet size. The costing is based on software development rates in SMAD chapter 20, converted to FY2004$. The heritage factors are assigned based on the approximate difficulty of changing the software. It is assumed that changes to the ARQ and routing protocols are more difficult than changes to the user data rate and data packet size.

The reversibly-connected reconfiguration transitions include the software changes to implement Automatic Repeat Request (ARQ) and routing protocol changes and to adjust the user data rate and packet size. Assumptions on the costing for these changes are included in Table 4.6. The costing is based on software development rates in SMAD chapter 20, converted to fiscal year 2004 dollars.[54] The calculations assume UNIX-C software and a 50/50 split between modifications to flight and ground software. The software modification is assumed to be made to the existing 6.3 thousand lines of code software. The heritage factors are assigned based on the approximate difficulty of changing the software. It is assumed that changes to ARQ and routing protocols are more difficult than changes to user data rate and the data packet size.

4.1.6 Estimating System Benefits

This section details how the system benefits (revenue) are estimated. The potential revenue is based on market research and the expected services that the constellation configurations can provide.

The estimation will take place in three parts: first, a combined market-traffic model will be presented, which describes the interactions between the market demand and the packet traffic statistics; a model for estimating the global subscriber demand will be described; and finally, the combined market-traffic-service triad model used to estimate the system revenue will be introduced.

Combined Market-Traffic Model

Since this case study introduces a new revenue estimation model, a detailed description of it is included here and in the following sections. The combined market-traffic model was first proposed in 2005.[7] The components are based on the work of Mohorcic et.al. and Kashitani.[70,71] The following sections extend this work to incorporate service modeling to create a market-traffic-service triad.

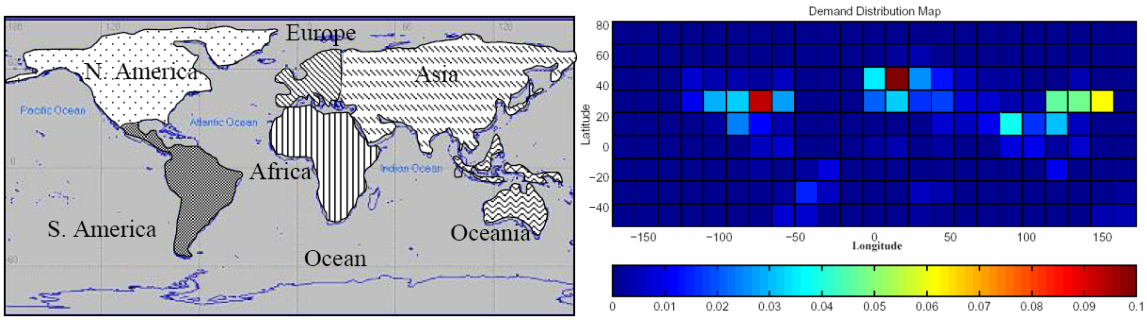
To reduce the computational expense of the simulation model, a grid of 30 28°-by-60° latitude-longitude rectangles were specified between ±70° latitude (the zone in which most of the world's population resides). Within each rectangle, the percentage of each geographical landmass (given as regions in Table 4.7 and Figure 4.5a) was estimated to the nearest 1/16 of a latitude-longitude rectangle. This information was used to scale the percentages in Table 4.7 so that the probability of destination given source in each source rectangle was properly normalized to account for the geographically-weighted probabilities.

To simplify routing calculations, it was assumed the traffic could be specified as independent Poisson processes. Each latitude-longitude rectangle can be assigned as a source node and a destination node, such that all traffic originating from a region is modeled as a Poisson process of a rate appropriate to the loading entering the source node for that region.

To determine the appropriate rate for each Poisson process, a model of the geographical market distribution is incorporated as follows.

Figure 4.5b represents a normalized matrix of the relative demand weightings distributed geographically around the world. The figure is based on work done by Kashitani,[71] in which it is assumed that a customer's willingness and ability to subscribe to a telecommunications service is dependent on their economic status and exposure to technology (modeled by the world map of Gross National Product adjusted by Purchasing Power Parity (GNP PPP). Furthermore, Kashitani assumes that the number of customers in an area willing to subscribe is likely also dependent on the population density.

If it is assumed that the normalized market distribution corresponds to the relative packet arrive rate weightings (λ^S) to each source node, the two models can be combined to enable transparent modeling of the effect of varying subscriber demand levels. The equation combining the two models is given by:



(a) Traffic Region Map (b) Demand Distribution Map

Figure 4.5: (a.) Geographical regions for traffic source-destination flow table and the (b.) market demand map. The demand map combines Gross National Product adjusted by purchasing power parity with the population densities to generate a normalized matrix of the relative demand weightings distributed geographically. Pictures taken from Underwood and Chan et.al., respectively.[7,8]

	Destination Region						
Source Region	N. America	S. America	Europe	Asia	Africa	Oceania	Ocean
N. America	83.3	2.9	3.9	3.9	2	2	2
S. America	6.8	79.4	6.8	2	2	1	2
Europe	3.9	2.9	82.5	3.9	2.9	1	2.9
Asia	4.9	1	4.9	81.3	2	3.9	2
Africa	4.9	2	6.9	4	80.2	1	1
Oceania	5	1	2	6.9	1	83.1	1
Ocean	35	5	25	30	2	2	1

Table 4.7: Total traffic flow between source and destination regions. Table given in percentages, or the conditional probability that a given data packet will travel to a given destination region from a given source region. Table from Underwood.[7]

$$\lambda^{SD}|_{Normalized} = \lambda^{S} P(S|D)$$

Another factor to be incorporated is the subscriber usage in terms of packets per second per subscriber. Several sources cite 125 minutes/month as a common usage factor.[25,16] It is necessary to convert this value to packets per second. At 1 Mbps for 100 byte packets, this value roughly translates to 3.44 packets/second (120 minutes/month), which is what the simulation assumes. Another revision of this model would be to refine the value to be a function of user data rate and packet size and to adjust according to the service provided (telephony, radio/TV, Internet, and etc).

The arrival rates ($\lambda^{SD}|_{Normalized}$) can now be scaled by the global subscriber demand to model the geographic distribution of the call origination as well as the distribution of calls from source to destination. The next section will describe how the global subscriber demand is modeled.

Modeling Global Subscriber Demand

The global subscriber demand for a given service can be modeled as the multiplication of three factors: the total global subscriber market for the service, the market fraction taken by satellite providers for that service and the business fraction achieved by the particular business within the satellite market for the service.

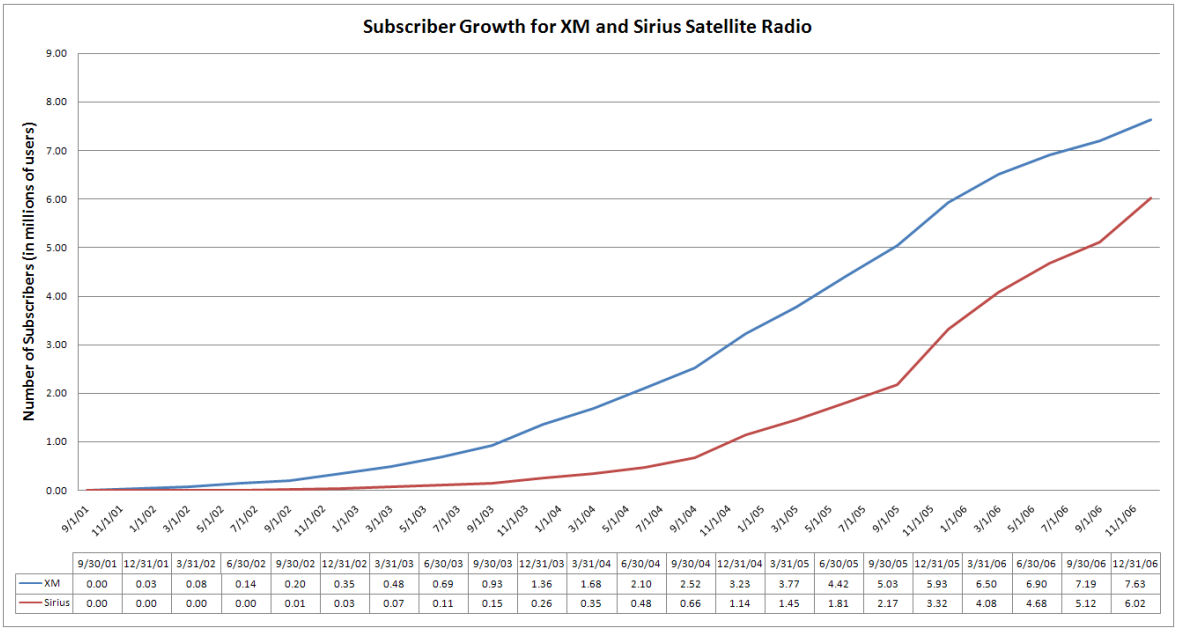
There are several models available for estimating the total global subscriber market for the various services satellites can provide. For example, Chang [72] models the potential communication satellite system market according to whether the system is low-bandwidth (user data rates less than 50 kbps) or high-bandwidth (user data rates greater than or equal to 50 kbps). The data is based off of the Globalstar system and its experience with penetrating the telephony market.

Sources of data for estimating subscriber demand include European Space Agency (ESA) Telecommunications. It is not uncommon for the satellite market fraction to be on the order of 10 percent to 20 percent of the total subscriber base, and the business fraction for satellites range from 1 percent to 21 percent. [73] It is difficult to estimate the standard deviation for the growth as the numbers tend to be for overall growth in a market segment, not broken down by a particular business (which can have a wildly different growth pattern from the overall pattern).

As will be shown in a following section, the legacy constellation provides a service similar to XM or Sirius Satellite radio with the addition of bulk data capability. For comparison, a time line of the subscriber demand for both is included in Figure 4.6. XM started service in September 2001. By December it had 27,733 subscribers. Sirius posted similar 4th quarter gains in its first year: 29,947 subscribers by December 2002. [10] Let it be assumed that the service in this case study also started at 27,733 subscribers.

The growth of both systems is quite remarkable. The average annual increment for Sirius and XM are 1.4985 million and 1.4763 million, respectively. [9] Compare to the total estimated annual increment for low-bandwidth satellite systems (that is, over all of the low-bandwidth services) of 2.9677 million users. In 2003, the total estimated subscriber level is 49.60 million users. [72] In the cases of XM and Sirius, subscriber growth remained strong, though there is always a chance of loss of subscribers. One thing is clear: it is very difficult to estimate the standard deviation of the market variations from the data.

A common model used in financial forecasting and for demand-level forecasting is known as Brow-



	9/30/01	12/31/01	3/31/02	6/30/02	9/30/02	12/31/02	3/31/03	6/30/03	9/30/03	12/31/03	3/31/04	6/30/04	9/30/04	12/31/04	3/31/05	6/30/05	9/30/05	12/31/05	3/31/06	6/30/06	9/30/06	12/31/06
XM	0.00	0.03	0.08	0.14	0.20	0.35	0.48	0.69	0.93	1.36	1.68	2.10	2.52	3.23	3.77	4.42	5.03	5.93	6.50	6.90	7.19	7.63
Sirius	0.00	0.00	0.00	0.00	0.01	0.03	0.07	0.11	0.15	0.26	0.35	0.48	0.66	1.14	1.45	1.81	2.17	3.32	4.08	4.68	5.12	6.02

Figure 4.6: XM and Sirius subscriber growth chart. Created from data obtained at the Satellite Standard Group and adapted from a similar graph on Wikipedia.[9,10]

nian Motion via the Weiner model. However, it is usually desired to evaluate the forecast over large intervals of time in order to minimize the computation time. Thus, it is often done using a binomial model. An approach is outlined in de Weck et.al.[74] in which expressions for u, d and p can be found as follows:

$$u = e^{\sigma\sqrt{\Delta t}}$$

$$d = \frac{1}{u}$$

$$p = \frac{e^{\mu\sqrt{\Delta t}} - d}{u - d}$$

Since the simulation evaluates over a number of experiments varying σ and scenarios randomly varying the up and down motion of the binomial model, it is not necessary to keep track of p. The histogram of results should indicate the distribution of outcomes.

The base σ is assumed to be 40%. This is the value used to calculate the 'actual' market scenario. In the forecasting model, this value is randomly varied from 20% to 60% over 15 experiments using a uniform distribution. Each experiment contains 10 scenarios in which the up and down motion of the model is randomly varied also according to a uniform distribution.

The forecast model sets the starting demand level to the last known 'actual' value.

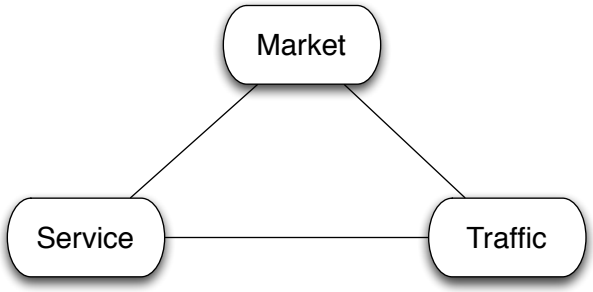


Figure 4.7: Market-Traffic-Service Triad

Revenue Estimation: the Market-Traffic-Service Triad

A key element of the architecture desirability and another contribution of the thesis is the market-traffic-service triad (see Figure 4.7). The market size and share, traffic load on the network and the services provided are interconnected in interesting ways. The coupling between these system aspects can affect its desirability from both the customer and decision maker point of view.

On the average, an increase in the number of users (market size and/or share) means an increased traffic load. This relationship is on the average because an estimate of the traffic load must be based on the average system usage by a customer. It's possible that the new customers will use the system less than the average, or more than the average.

The types of services provided also impacts the traffic load (i.e., number of packets). Some applications generate greater traffic load than others (See Appendix D).

Why is the traffic load important? Increased traffic load may cause congestion which can increase the round-trip delay, jitter and loss, all of which are key quality of service metrics that affect the types of services that can be provided according to customer expectations. One could argue that a drop in these quality-of-service metrics does not necessarily mean that the applications can no longer be supported, but it may make the service unusable or irritating from a customer point of view. It's important to remember that the satellite system must compete with terrestrial systems, so if the terrestrial systems provide services with these minimum quality-of-service standards, then the satellite systems must do the same to compete successfully. If using the system becomes irritating to customers, some will migrate to competing systems.

Generally, this affect only occurs as the system approaches saturation, which would indicate that an architecture transition may be necessary in order to maintain customer levels.

If architecture transitions are not undertaken, there are a few things that can happen. First, the decision-makers' may choose to do nothing. In all likelihood, many customers will migrate to competing systems in order to maintain the services and applications they were promised and had been paying for. If enough customers leave, then an interesting thing will happen. The traffic load will decrease which will act to improve the quality of service back toward the previous levels. But now there are fewer paying customers and it will be hard to attract new customers, and if enough new customers are brought online, then the cycle may repeat.

The second thing that can happen is that the decision-makers' may choose to simply switch the business-area focus of the system to try to capture new markets. Doing so will alienate some customers because they may lose some of the services promised to them, or at the very least, they will experience a drop in the quality of service of those services. But now, new markets may be tapped and it will likely be some time before the system hits the new saturation point. Naturally, one would expect that the new markets will not produce the level of revenue per customer as before since the services offered will be less than before, but creative marketing can increase the demand level to offset the necessary drop in monthly usage fees.

The MATLAB simulation used for this case study captures the steady-state network behavior for the constellation at an instant in time as a function of the demand level. The demand model combines the market and traffic aspects of the triad, producing the global geographic distribution of the arrival rates of packets, scaled by an estimated customer usage factor.

The steady-state network behavior includes evaluations of key quality of service metrics (average round-trip delay, jitter data loss). A combination of these metrics with the average user data rate and the average user packet size can be mapped to the service regions bounded by minimum quality of service standards (see Table 4.8 on page 204). Combinations of these service regions can be mapped to the business areas shown in Table 4.9 on page 205 (e.g., provides the services required for XM radio with weather and navigation features). These business areas can be compared to real-world terrestrial and current successful satellite systems to estimate competitive monthly usage fees.

A radar map showing the overlapping regions of the service regions is shown in Figure 4.8 on page 204. These regions correspond to the potential business areas. Relevant business-area regions have been noted on the radar map.

This case study has introduced a mechanism to estimate system revenue as a function of the number of subscribers and their geographic distribution around the world, the average estimated subscriber

Service Regions							
Service Region	**Description**	**Data Rate** [kbps]	**Packet Size** [Byte]	**Loss** [% data loss]	**Average RTD** [sec]	**Jitter** [sec]	**Value** [$]
1	Minimum Telephony (Voice over Internet Protocol (VoIP))	≥ 10	≤ 10	≤ 1	≤ 0.3	≤ 0.06	$5
2	Excellent Telephony (VoIP)	≥ 100	≤ 100	≤ 1	≤ 0.3	≤ 0.06	$10
3	Streaming Radio and Streaming Low-quality Video	≥ 100	≤ 100	≤ 5	≤ 10	–	$15
4	High-quality Streaming Video	≥ 1000	≤ 100	≤ 5	≤ 10	–	$20
5	Interactive Streaming Video	≥ 1000	≤ 1000	≤ 1	≤ 0.3	≤ 0.06	$25
6	Interactive Data	≥ 10	≤ 100	≤ 3	≤ 0.3	≤ 0.06	$25
7	Transactional Data	≥ 100	≤ 1000	≤ 5	≤ 0.5	≤ 0.1	$20
8	Bulk Data	≥ 10	≤ 1000	≤ 20	≤ 60	≤ 5	$10
9	Best-effort Data	≥ 100	≤ 1000	≤ 10	≤ 5	≤ 0.5	$10

Table 4.8: Service regions. The table includes the data rate, packet size and quality-of-service performance characteristics (loss, average round-trip delay and jitter) required to achieve each of the service regions. The approximate market value of each service is noted. (Note: RTD is short for round-trip delay.) This table is based on the table in Appendix D.

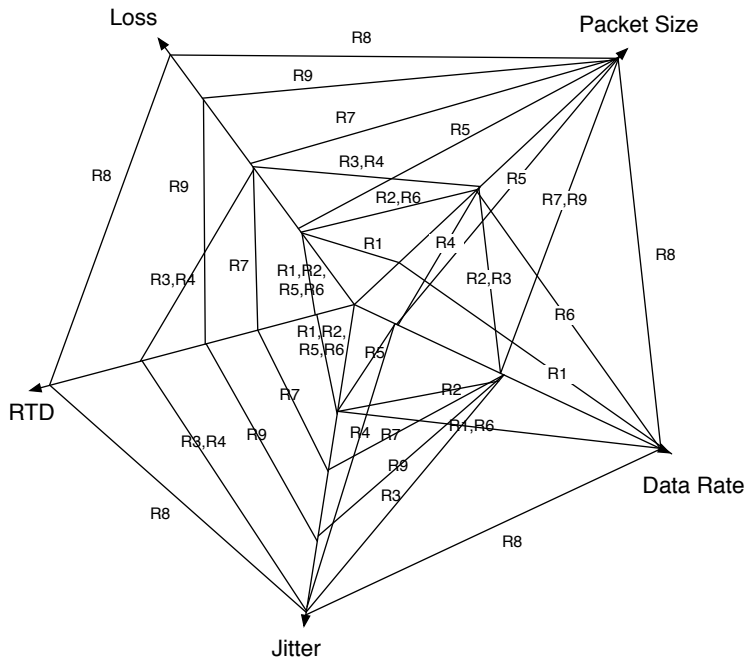


Figure 4.8: Radar map of the overlapping service region areas as bounded by the limit of the data rate, packet size and quality-of-service performance characteristics given in Table 4.8. These areas can be mapped to the achievable business areas.

Business Areas			
Business Area Category	**Service Region Categories**	**Competitive Cost/user/month**	**Example**
1	1, 8	$15	Similar to Onstar
2	8	$10	
3	1, 2, 3, 7, 8	$60	Low-end Wireless Plan
4	3, 7, 8, 9	$55	
5	2, 3, 6, 7, 8, 9	$90	High-end Wireless Plan
6	1, 2, 3, 4, 7, 8	$80	Mid-range Wireless Plan
7	2, 3, 4, 5, 6, 7, 8, 9	$135	Voice and Data Plan
8	8, 9	$20	
9	3, 7, 8	$45	
10	3, 4, 7, 8	$65	
11	3, 4, 7, 8, 9	$75	
12	3, 8, 9	$35	
13	3, 8	$25	
14	3, 4, 8	$45	
15	7, 8, 9	$40	
16	3, 4, 8, 9	$55	
17	3, 4, 6, 7, 8, 9	$100	
18	3	$15	XM Radio
19	3,4	$35	

Table 4.9: Business Areas with Approximate Competitive Monthly Subscription Fees. Business Areas without a real world example estimate the appropriate monthly usage fees, otherwise, the usage fees are comparable to the example listed. (Note: Onstar uses terrestrial network for telecommunications and GPS for nav features.)

usage, and the resulting competitive services that can be provided on the basis of quality of service standards.

4.1.7 Options

There are a large number of feasible options available to the decision-makers, enough to make computation of the entropy metric difficult (see Table 4.10 on page 206). There are numerous ways to artificially limit the option space, which can speed up the process considerably, reduce confusing and unnecessary clutter on the Advantage Map and eliminate options which perform poorly relative to the legacy architecture.

Reducing the Option Space

The legacy architecture brings in a competitive usage fee of $25/month per user on average. Thus, there is no reason to include any architecture that will not bring a revenue stream greater than this amount. The option space can now be reduced to that shown in Table 4.11 on page 207.

This reduced set of options can be converted to an architecture table similar to Table 4.12 on page 207. This table shows there are 31 feasible architectures in 3 constellations. This is still a rather large set of architecture options to consider.

Business Areas			Monthly User Fee						
1	2	3	1	2	3	Ruser (kbps)	RP	ARQ	PS (Bytes)
	2	1		10	15	10	1	1	10
	2	2		10	10	10	1	1	100
		2			10	10	1	2	100
	2	2		10	10	10	1	3	10
	2	2		10	10	10	1	3	100
2	2	2	10	10	10	10	2	1	10
	2	2		10	10	10	2	3	10
2	2	2	10	10	10	10	3	1	10
	2	2		10	10	10	3	3	10
	9	3		45	60	100	1	1	10
13	12	12	25	34	34	100	1	1	100
		2			10	100	1	1	1000
	2	2		10	10	100	1	3	10
	2	12		10	35	100	1	3	100
13	13	3	25	25	60	100	2	1	10
13	13	13	25	25	25	100	2	1	100
	2	9		10	45	100	2	3	10
		2			10	100	2	3	100
13	13	3	25	25	60	100	3	1	10
13	13	13	25	25	25	100	3	1	100
	2	9		10	45	100	3	3	10
		2			10	100	3	3	100
19	10	6	35	65	80	1000	1	1	10
16	11	7	55	75	135	1000	1	1	100
	2	8		10	20	1000	1	1	1000
	2	2		10	10	1000	1	3	10
	2	11		10	75	1000	1	3	100
		2			10	1000	1	3	1000
14	14	6	45	45	80	1000	2	1	10
16	16	11	55	55	75	1000	2	1	100
	2	10		10	65	1000	2	3	10
		8			20	1000	2	3	100
14	14	6	45	45	80	1000	3	1	10
16	16	11	55	55	75	1000	3	1	100
	2	10		10	65	1000	3	3	10
		8			20	1000	3	3	100

Table 4.10: Business area and monthly user fee trends for each architecture instantiation, assuming subscriber load of 50,000 subscribers. Only includes architectures that performed sufficiently well to return a business area.

Business Areas			Monthly User Fee						
1	2	3	1	2	3	Ruser (kbps)	RP	ARQ	PS (Bytes)
	9	3		45	60	100	1	1	10
	12	12		35	35	100	1	1	100
		12			35	100	1	3	100
		3			60	100	2	1	10
13			25			100	2	1	100
		9			45	100	2	3	10
		3			60	100	3	1	10
		9			45	100	3	3	10
19	10	6	35	65	80	1000	1	1	10
16	11	7	55	75	135	1000	1	1	100
		11			75	1000	1	3	100
14	14	6	45	45	80	1000	2	1	10
16	16	11	55	55	75	1000	2	1	100
		10			65	1000	2	3	10
14	14	6	45	45	80	1000	3	1	10
16	16	11	55	55	75	1000	3	1	100
		10			65	1000	3	3	10

Table 4.11: Business area and monthly user fee trends for each architecture instantiation, assuming subscriber load of 50,000 subscribers. The data is filtered to ensure a usage fee greater than the current fee of $25 per month).

	Decision Variables					50,000 Subscribers	
Architecture	Constellation	Ruser (kbps)	RP	ARQ	PS (Bytes)	Business Area	Monthly User Fee
1 (Legacy)	1	100	2	1	100	13	$25
2	1	1000	1	1	10	19	$35
3	1	1000	1	1	100	16	$55
4	1	1000	2	1	10	14	$45
5	1	1000	2	1	100	16	$55
6	1	1000	3	1	10	14	$45
7	1	1000	3	1	100	16	$55
8	2	100	1	1	10	9	$45
9	2	100	1	1	100	12	$35
10	2	1000	1	1	10	10	$65
11	2	1000	1	1	100	11	$75
12	2	1000	2	1	10	14	$45
13	2	1000	2	1	100	16	$55
14	2	1000	3	1	10	14	$45
15	2	1000	3	1	100	16	$55
16	3	100	1	1	10	3	$60
17	3	100	1	1	100	12	$35
18	3	100	1	3	100	12	$35
19	3	100	2	1	10	3	$60
20	3	100	2	3	10	9	$45
21	3	100	3	1	10	3	$60
22	3	100	3	3	10	9	$45
23	3	1000	1	1	10	6	$80
24	3	1000	1	1	100	7	$135
25	3	1000	1	3	100	11	$75
26	3	1000	2	1	10	6	$80
27	3	1000	2	1	100	11	$75
28	3	1000	2	3	10	10	$65
29	3	1000	3	1	10	6	$80
30	3	1000	3	1	100	11	$75
31	3	1000	3	3	10	10	$65

Table 4.12: Architecture table for the reduced set of options.

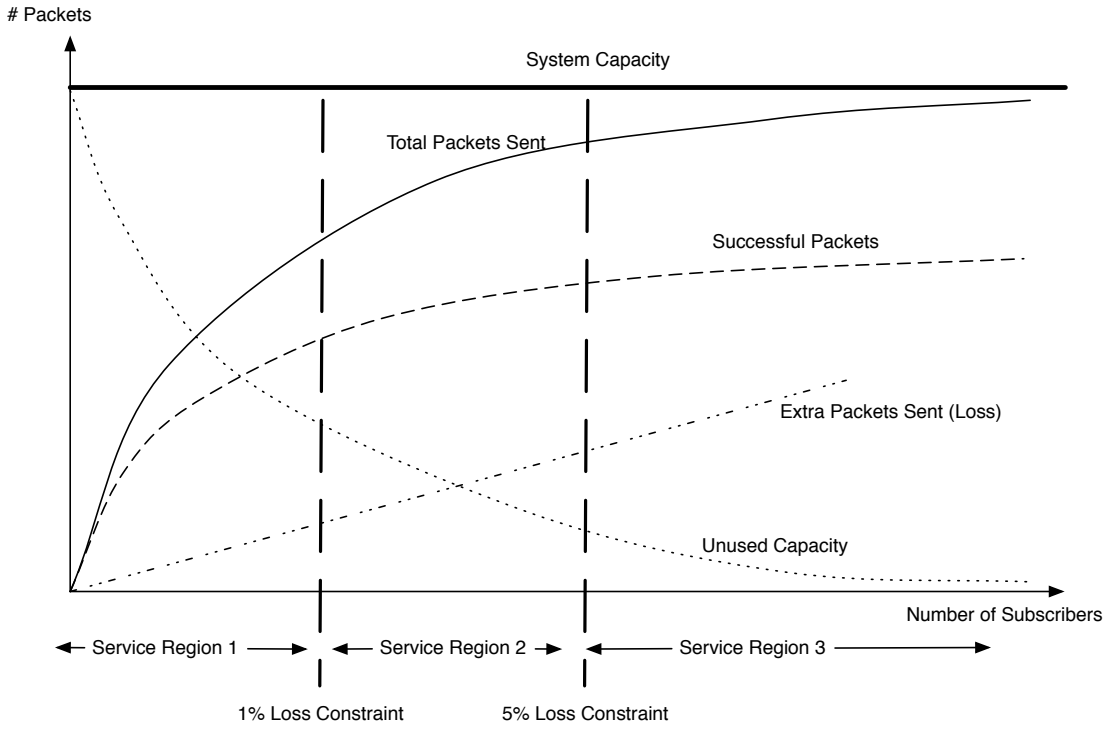


Figure 4.9: Saturation graph.

Estimation of System Response

Running the energy and entropy calculations can require a significant amount of computation time. It is sometimes possible to model and sample the system response to uncertainties ahead of time in order to further eliminate options and to speed up the metric calculation process.

For this case study, the uncertainty takes the form of market demand, which drives the traffic load on the system and thus the service regions and business areas achievable by the system. The business area dictates the competitive usage fee that can be obtained from each subscriber, thus, the system response (business area) to the uncertainty (market demand) drives the revenue the system can bring in.

But, sometimes the system can become saturated, and the system can start losing achievable service regions (see Figure 4.9 on page 208). When this happens, the usage fee drops to the next achievable service region. This models two things: first, the decision makers could choose to switch business areas in order to target a new market and thus the achievable revenue per subscriber drops. Second, if the decision makers choose to keep the business area, with a reduced quality of service, it is likely that the system would lose subscribers which is likewise reflected by a drop in total revenue. In

System Response			
Architecture	Business Area Response	Demand Switch	Monthly Usage Fee Switch
1 (Legacy)	$13 \rightarrow 18 \rightarrow 0$	$1.70M, 6.45M$	$\$25 \rightarrow \$15 \rightarrow \$0$
2	19	-	\$35
3	16	-	\$55
4	14	-	\$45
5	16	-	\$55
6	14	-	\$45
7	16	-	\$55
8	9	-	\$45
9	12	-	\$35
10	10	-	\$65
11	$11 \rightarrow 16$	2.35M	$\$75 \rightarrow \$55$
12	14	-	\$45
13	16	-	\$55
14	14	-	\$45
15	16	-	\$55
16	$3 \rightarrow 9$	1.10M	$\$60 \rightarrow \$45$
17	$12 \rightarrow 13$	4.50M	$\$35 \rightarrow \$25$
18	$12 \rightarrow 8 \rightarrow 2 \rightarrow 0$	$100K, 400K, 4.9M$	$\$35 \rightarrow \$20 \rightarrow \$10 \rightarrow \$0$
19	$3 \rightarrow 9 \rightarrow 13$	$100K, 1.55M$	$\$60 \rightarrow \$45 \rightarrow \$25$
20	$9 \rightarrow 13 \rightarrow 2 \rightarrow 0$	$100K, 300K, 5.7M$	$\$45 \rightarrow \$25 \rightarrow \$10 \rightarrow \$0$
21	$3 \rightarrow 9 \rightarrow 13$	$100K, 1.65M$	$\$60 \rightarrow \$45 \rightarrow \$25$
22	$9 \rightarrow 13 \rightarrow 2 \rightarrow 0$	$100K, 350K, 5.85M$	$\$45 \rightarrow \$25 \rightarrow \$10 \rightarrow \$0$
23	6	-	\$80
24	$7 \rightarrow 11$	200K	$\$135 \rightarrow \$75$
25	$11 \rightarrow 8 \rightarrow 2$	$150K, 2.2M$	$\$75 \rightarrow \$25 \rightarrow \$10$
26	6	-	\$80
27	$11 \rightarrow 16$	450K	$\$75 \rightarrow \$55$
28	$10 \rightarrow 2$	1.65M	$\$65 \rightarrow \$10$
29	6	-	\$80
30	$11 \rightarrow 16$	650K	$\$75 \rightarrow \$55$
31	$10 \rightarrow 2$	1.8M	$\$65 \rightarrow \$10$

Table 4.13: The system response specified using the architecture saturation transition points. The demand switch is noted as the first sampled level of demand that switches the achievable business area.

either case, the total revenue is reduced and it can be assumed to be modeled by the business area of the system and the corresponding competitive usage fee.

The simulation model is run for each revenue-generating architecture (the architectures that return a business area for Iridium subscriber level, 60,000 subscribers) over a wide-range of subscriber levels (50,000 to 20,000,000 at intervals of 50,000). The transition points were noted and included in Table 4.13.

Each of the architectures that could not do better than the legacy architecture in Constellation 1 were immediately eliminated from consideration. Using similar logic, it makes sense to eliminate all ar-

Software Commonality					
Architecture	Ruser	RP	ARQ	PS	Usage Fee
1 (Legacy)	100	2	1	100	\$25 → \$15 → \$0
2	1000	1	1	10	\$35
3	1000	1	1	100	\$55
4	1000	2	1	10	\$45
5	1000	2	1	100	\$55
6	1000	3	1	10	\$45
7	1000	3	1	100	\$55
10	1000	1	1	10	\$65
11	1000	1	1	100	\$75 → \$55
23	1000	1	1	10	\$80
24	1000	1	1	100	\$135 → \$75
26	1000	2	1	10	\$80
29	1000	3	1	10	\$80

Table 4.14: Software commonality of the remaining options.

chitectures in Constellation 2 that cannot do at least as well as the best architecture in Constellation 1 (usage fee of \$55). This eliminates all but Architectures 10 and 11 from consideration in Constellation 2. Similarly, all but Architectures 23, 24, 26 and 29 can be eliminated from consideration in Constellation 3.

There are thirteen architecture options remaining. It may be possible to further reduce this set by considering the decision commonality between them. Any transition that requires more software changes to achieve the same response is dominated by the transitions that require fewer software changes and thus incur less energy by transitioning. It should therefore be possible to eliminate options that repeat a usage fee and are dominated by another option. The software commonality of the remaining architectures is shown in Table 4.14.

The architectures of interest to keep are 2, 10, 11 and 24, since they all have non-repeating usage fees (note: the ones that transition are interesting, even if they do transition to a usage fee that is achievable by a non-transitioning architecture). It should be clear from this set that Architectures 6, 7, and 29 can be automatically eliminated due to RP 3, reducing the set to 10 architectures.

Ten architecture options is a sufficiently reduced set for relatively fast computation of the metrics. However, to ensure simple illustration, it would be best to further eliminate the number of architectures down to at most seven. Architectures 2, 10, 11 and 24 share RP 1. The architectures with

Final Set of Options with System Response				
Architecture	Business Area Response	Demand Switch (start of new)	Monthly Usage Fee Switch	Assigned Color
1 (Legacy)	$13 \rightarrow 18 \rightarrow 0$	$1.70M, 6.45M$	$\$25 \rightarrow \$15 \rightarrow \$0$	Blue
2	19	-	\$35	Green
3	16	-	\$55	Red
4	10	-	\$65	Cyan
5	$11 \rightarrow 16$	2.35M	$\$75 \rightarrow \$55$	Magenta
6	6	-	\$80	Yellow
7	$7 \rightarrow 11$	200K	$\$135 \rightarrow \$75$	Black

Table 4.15: Final set of options with system response.

RP 2 (Architectures 4, 5 and 26) share comparable usage fees to the architectures with RP 1. It does not make sense to incur a transition cost to achieve the same usage fee. For this reason, we can further eliminate Architectures 4, 5 and 26. (Thus, in the final option space, once the system is transitioned away from the legacy configuration, the only variables that change are the constellation specification and the packet size.)

Final Set of Options

The final set of options to be considered are documented in Table 4.15, including a notation of the assigned colors for use in the Strategic Advantage and Exposure maps. The corresponding final option network is shown in Figure 4.10 on page 212.

4.2 Position

This section describes how the the system position is described and evaluated for this case study. The current energy of the system is found, the method used to estimate the terminal entropy values is discussed, the weightings chosen for use in the case study are noted, and the initial Strategic Advantage Maps and Iceberg exposures are presented and explained.

4.2.1 Current Energy

The current energy can be found by summing up the actual expenses incurred over the lifetime of the system to-date as well as the actual revenues gained.

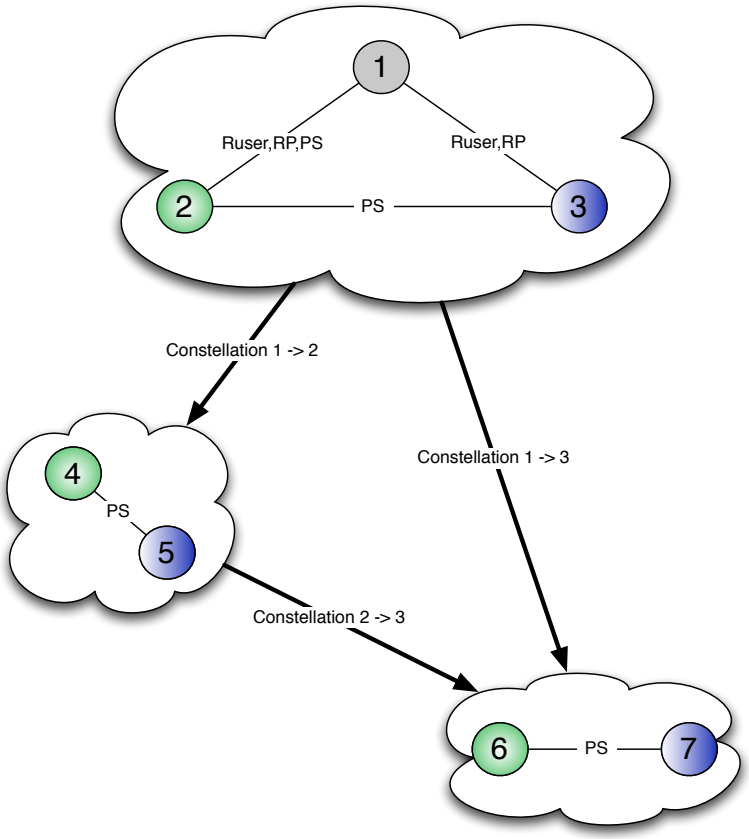


Figure 4.10: Final option network corresponding to Table 4.15.

For this case study, let it be assumed that the revenues and expenditures occurred as shown in Figure 4.11, and that they are all in fiscal 2004 dollars. The initial development cost was $1.5152 billion. The operating expenses varied a bit, from $55.314 million to $65.053 million. The total expenses in fiscal 2004 dollars is $1.8212 billion.

The total revenues generated are noted in Figure 4.11. The system has collected an average of $25/month from each customer. The revenues have steadily increased over the first five years, from $12.812 million to $81.273 million, following the "actual" demand trends in Figure 4.12. The total revenue collected in fiscal 2004 dollars is $191.933 million, nowhere near the amount spent, even when considering the total operating costs of $306 million.

This gives the system an initial energy of 9.4887.

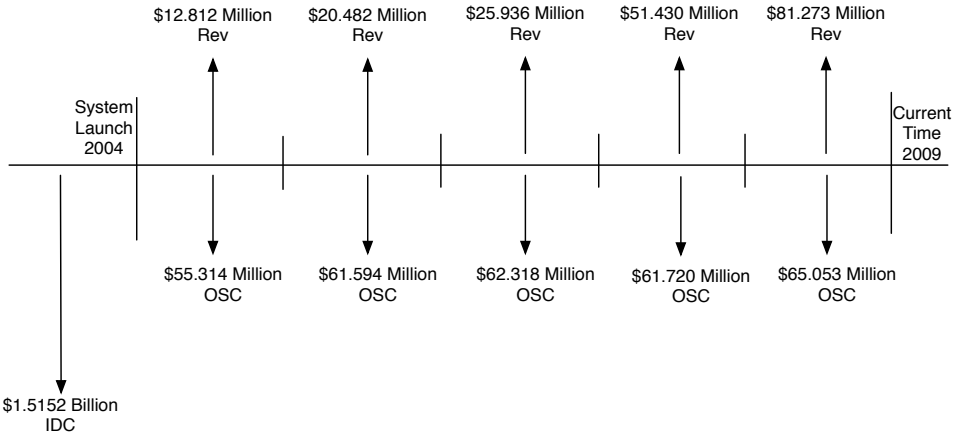


Figure 4.11: Timeline of historical expenses and revenues, in fiscal 2004 dollars.

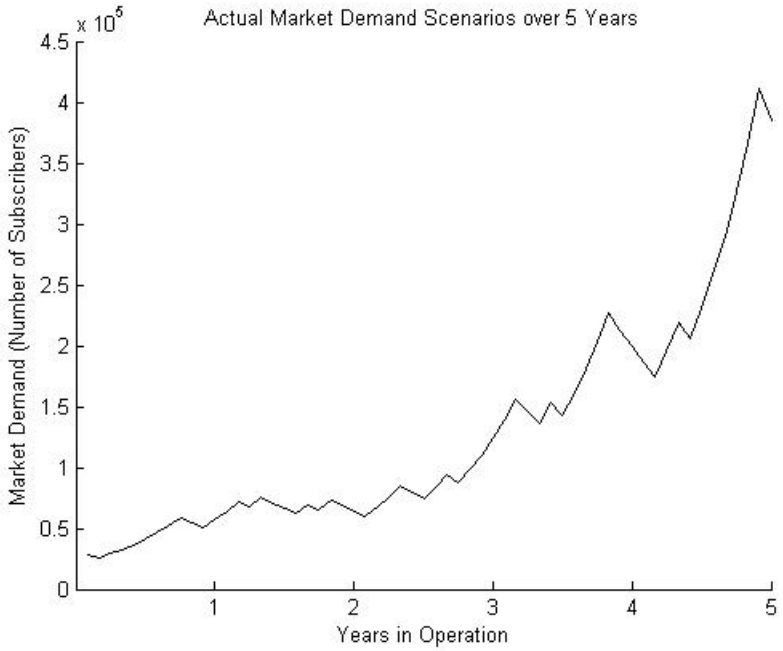


Figure 4.12: Actual market demand for the five year history.

4.2.2 Estimation of Terminal Values

The entropy-based formulation requires that terminal values be set to weight the horizon. For this case study, the horizon is set to be the end-of-life of the current hardware. Thus, there are a finite set of actions that the decision makers can make at the horizon, based on the evolution rules: retire the constellation, rip out the constellation and build new, or allow the constellation to be reduced to the set of newer satellites that have been launched in the interim. It is assumed, for simplicity, that the retirement cost is the same as the annual maintenance cost.

If the architecture is retired, its long-term (call it "retirement") entropy is assumed to be zero. However, there is an energy adjustment due to the retirement itself. This calculation is identical to the retirement energy accounting discussed in the previous chapter.

If the architecture is ripped out and built new, it is assumed that the current system is retired and the "optimal" design in the current set of options is launched in full at the horizon. The terminal entropy is then estimated by calculating the entropy associated with operating this new system for another 15 years at the current operating cost and current revenue. At the end of this second 15-year period, the system is assumed to be retired with that long-term entropy set to zero as in the retirement case.

If the architecture is a mix of the original constellation and one or more new constellations, then the constellation at the horizon is reduced when the original constellation must be retired. At this point, the reduced set operates until the satellites reach their retirement life, and these operate until the second reduced set must be retired if there are two sets of new constellations. The cost and revenue streams associated with the reduced constellations were estimated, and the associated entropies calculated, in a manner similar to the rip out and build new case.

4.2.3 Weightings

For the purposes of this case study, all of the β weights are set to unity, implying that there is no preference for the different terms of the equation.

The α and γ debias factors were chosen in a preliminary attempt to calibrate the entropy formulation such that the initial strategic map made sense in the context of the perspective described earlier in this chapter.

The α weight for the configuration entropy is assumed to be of the form:

$$\alpha = \frac{1}{-\log\left(\min \frac{1}{C(k)}\right)}$$

The debias γ factor implemented in this case study is intended to reduce the biasing effect of the number of option paths for each immediately available transition. It is currently assumed to be the inverse of the outdegree (number of available transitions) for the architecture that exists at the evaluated time along each option path.

These factors are not ideal debiasing values, and future work should investigate methods to determine more appropriate metrics.

4.2.4 Initial Strategic Advantage Map and Iceberg Exposure

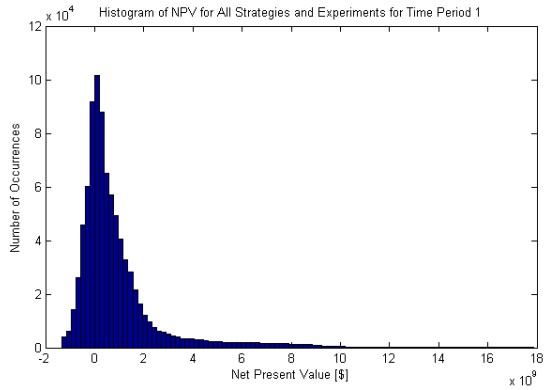
First, the initial change exposures for this case study will be presented and discussed. This section will also describe how to read the graphs in detail.

For reference, the architectural options that are available for immediate transition are separated by line color. The final architecture in the option path over the indicated number of transitions is identified by the point color. The region of possible outcomes for each immediate transition over the evaluated number of time periods is represented by the convex hulls traced by the colored lines. These regions form as a consequence of the monte carlo sampling of the demand uncertainty.

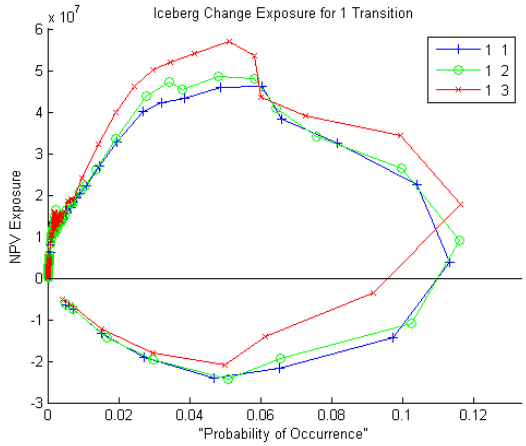
The initial histogram of possible net present value as projected over 10 periods, the corresponding Iceberg Exposure, and the initial Strategic Advantage Map looking out one time period are shown in Figure 4.13 on page 216.

The histogram of net present value and the Iceberg Exposure graph indicate a much greater potential for achieving a return on the investment than for losing money on the system. The Iceberg Exposure also demonstrates that transitioning to Architecture 3 appears to more desirable than the other options.

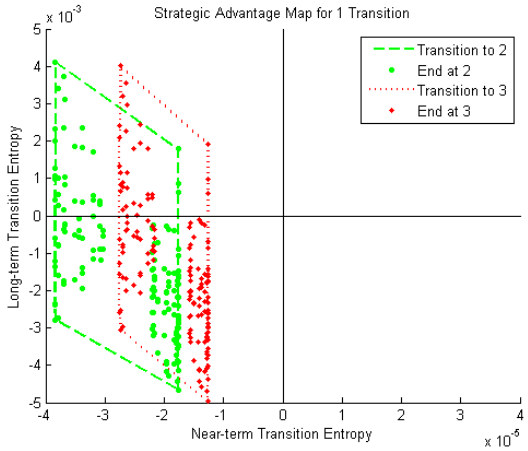
The Strategic Advantage Map plots the near-term and long-term position as viewed on the basis of a single transition (the immediate transition option). The fact that both immediate options (Architecture 2 and 3 only due to the evolution rules) have worse near-term position relative to the legacy is because there is a cost to transition without the benefit of the new (higher) usage fee. It is evident from the pattern that it costs more to transition to Architecture 2 than it does to transition to Architecture 3. Both transitions require changing the user data rate and the routing protocol. However, Architecture 2 must also adjust the packet size.



(a) Net Present Value



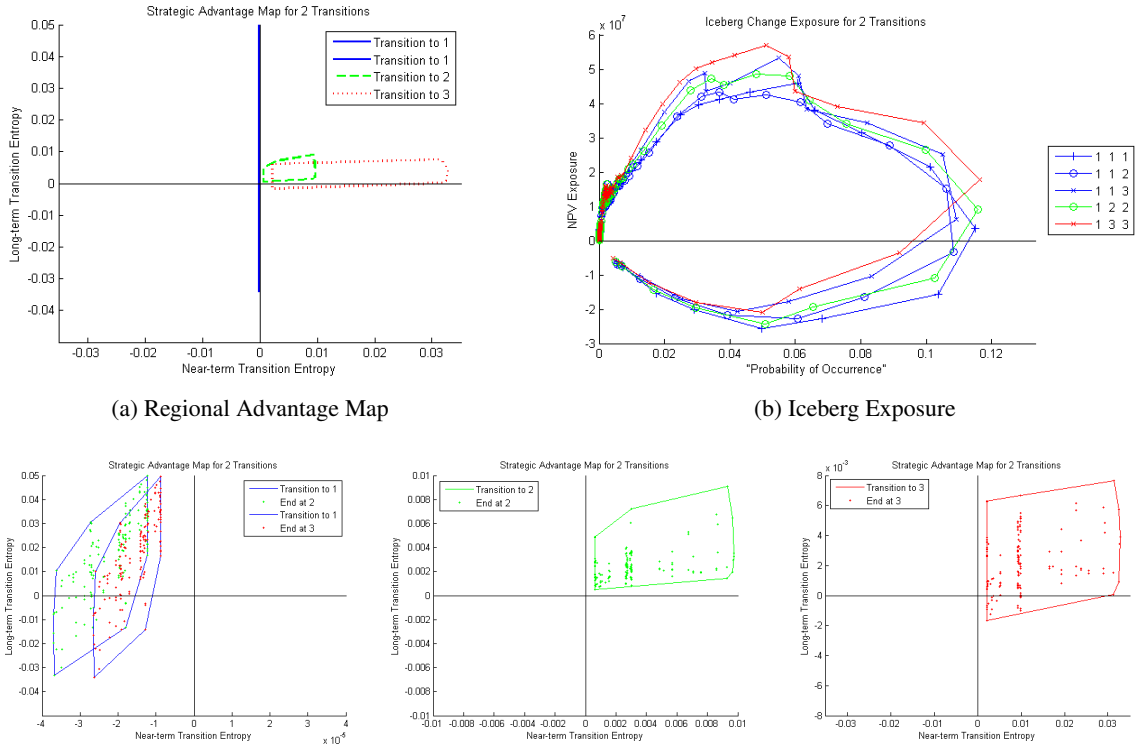
(b) Iceberg Exposure: One Transition



(c) Advantage Map: One Transition

Figure 4.13: Looking out one transition for Time Period 1: the (a) histogram of possible Net Present Value as projected over 10 periods, the corresponding (b) Iceberg Exposure and the (c) Strategic Advantage Map colored by transition.

The regions associated with the two immediate transitions fall substantially within Quadrant III, making them appear, on first glance, as poor choices. However, it is important to remember Evolution Rule 1 which states that, following a transition, an architecture must be monitored for at least one year before transitioning again. For this reason, the configuration entropy term is biased towards the legacy system, though the differences between the transition options and the legacy are very minimal (here, on the order of 1e-5). Thus, it is important to view more than just the Advantage Map for the first transition. Based on this information, transitioning to Architecture 3 is looking like a possibility, but there is still some chance of it being dominated by the legacy. To gain more information on the true strategic advantage of the relative options, the patterns in the next time period must be evaluated.



(a) Regional Advantage Map

(b) Iceberg Exposure

(c) Architecture 1

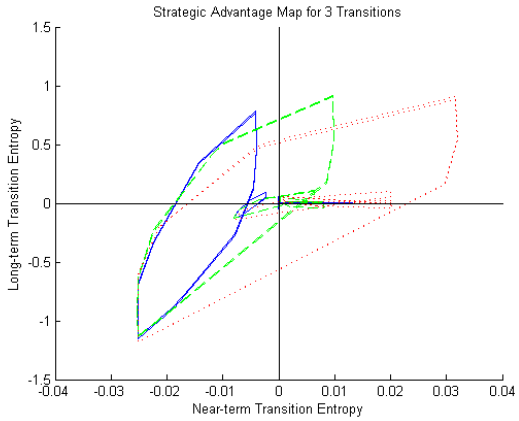
(d) Architecture 2

(e) Architecture 3

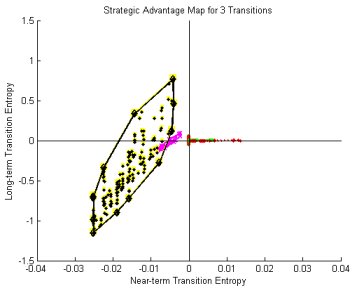
Figure 4.14: Looking out two transitions for Time Period 1: the (a) Strategic Advantage Map colored by immediate transition only (regional) and the (b) corresponding Iceberg Exposure over two transitions. The Strategic Advantage Maps for Time Period 1 looking out two transitions and filtered by whether the immediate transition is to (c) Architecture 1, (d) Architecture 2, or (e) Architecture 3.

The Advantage Map that looks out two transitions in Figure 4.14 on page 217 paints a very different story, largely due to the effect of the new (higher) usage fees and the elimination of the effect of Evolution Rule 1. The first figure shows the pattern of regions for all three architecture options on the same scale. The second figure is the corresponding Iceberg Exposure looking out two time periods. The lower three graphs separate the regional map into the three immediate transitions and zooms in on each to capture additional information, especially the separation of regions for maintaining Architecture 1 as is.

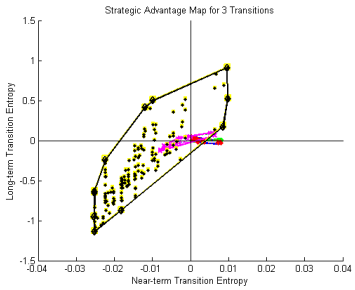
The Regional Advantage Map pattern along the near-term position axis is clearly due to the spread of the usage fees ($25 for the legacy architecture, $35 for Architecture 2 in green, and $55 for Architecture 3 in red). The regions hugging the long-term position axis are hard to see on this graph, but a close examination (see sub-figure (c)) will show two blue-lined regions to the left of



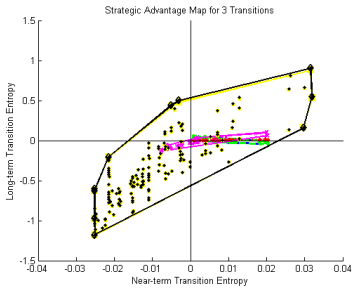
(a) Strategic Advantage Map



(b) Architecture 1



(c) Architecture 2



(d) Architecture 3

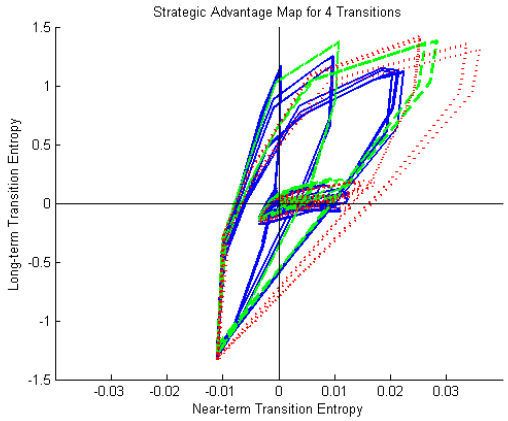
Figure 4.15: The (a) Strategic Advantage Map for Time Period 1 looking out three transitions colored by the immediate transition, and the Strategic Advantage Maps for Time Period 1 (colored by the architecture reached in four transitions) looking out three transitions and filtered by whether the immediate transition is to (b) Architecture 1, (c) Architecture 2, or (d) Architecture 3.

the axis, one with green dots and the other with red. These regions represent the options to stay in the legacy architecture for the immediate transition, then to transition to either Architecture 2 or Architecture 3 in the second transition period. The explanation for the pattern is the same as for the Advantage Map for the first transition. These near-axis striations appear to have a long-term position advantage, but this is due to the fact that the next transition and new-state terms in the entropy formulation strongly favor these option paths. In these two cases, the new usage fees have just kicked in. Thus, the energy drop from one state to the next is more significant than for the legacy or for the immediate transition to Architecture 2 or Architecture 3 since the next transition and new-state entropies for these option paths have stabilized (same usage fee, higher fees and thus improved revenue and energy, but not to the significant extent as the 1-1-2 or 1-1-3 paths). Based on this map, Architecture 3 is now the leading possibility for best choice. However, it is a good idea to look at least one more period out to get a better sense, particularly since the effect of some of the

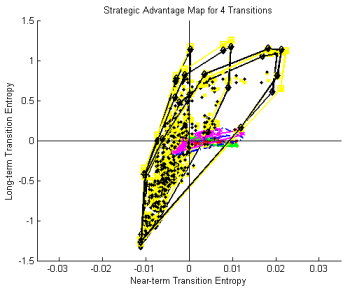
evolution rules have yet to kick in.

Looking out three transitions, as shown in Figure 4.15 on page 218, the effect of being able to transition to Constellation 2 (cyan and magenta) and 3 (yellow and black) becomes apparent. The top Strategic Advantage Map shows the pattern of results colored by immediate transition (Architecture 1, 2 or 3). The bottom graphs filter these results by the immediate transition and are colored by the architecture achieved in three transitions. Evolution Rule 2 states that there is a lag of three time periods before being able to reconfigure the constellation because of the need to manufacture the satellites and to make launch and operational preparations. Thus, the effect of these lower altitude options does not become apparent until the third transition. This produces some interesting results. Even though it is the first period where it is possible for the system to be within Constellation 2 or Constellation 3 (the transition costs are accounted for but not the new usage fees yet), there are some instances of superior performance relative to the 1-1-1-1 path. This demonstrates the huge difference in outcomes between the services offered (and the corresponding usage fees), even with some fairly pricey reconfiguration costs. Of course, there are plenty of instances that fall within the least desirable Quadrant III. However, this is likely due to not yet accounting for the increase in usage fees for these new architectural possibilities. All of the immediate transitions achieve roughly the same maximum long-term position from this view and meet at roughly the same place for the minimum long-term position, so using long-term position as a measure of desirable transition is not very helpful. Immediately transitioning to Architecture 3 followed by transitions to Constellation 3 have the best maximum near-term position. Transitioning to Constellation 2 is a close second, but the long-term position isn't nearly as desirable. Remaining in Architecture 3 past this period is strongly dominated by transitioning to Constellation 2 or Constellation 3. Remaining in Constellation 1 and transitioning within is even more dominated. This map confirms the earlier suspicion that the immediate transition to Architecture 3 is the best choice.

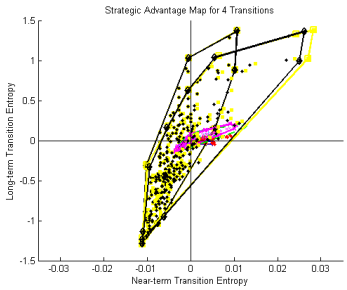
The map that looks out four transitions, shown in Figure 4.16 on page 220, continues to suggest that an immediate transition to Architecture 3 is the best choice. Again, the top Strategic Advantage Map shows regions of immediate transitions, and the bottom graphs are separated by immediate transition and colored by the architecture reached in four transitions. There is still some possibility of falling into Quadrant III but the pattern has moved significantly closer to the origin relative to the three-transition maps, suggesting a contraction and possible future migration into the other quadrants. Again, Architecture 3 dominates the immediate transition possibilities, with the eventual



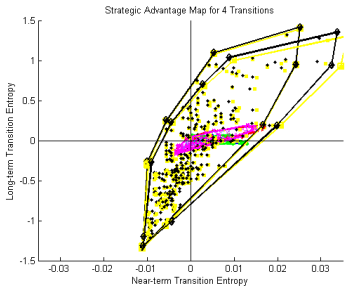
(a) Strategic Advantage Map



(b) Architecture 1



(c) Architecture 2



(d) Architecture 3

Figure 4.16: The (a) Strategic Advantage Map for Time Period 1 looking out four transitions colored only by the immediate transition, and the Strategic Advantage Maps for Time Period 1 looking out four transitions and filtered by whether the immediate transition is to (b) Architecture 1, (c) Architecture 2, or (d) Architecture 3. These graphs are colored by the architecture reached in four transitions.

transition to Constellation 3 continuing to be a likely strategic decision.

The recommended strategy based on these maps would be to immediately transition to Architecture 3 and begin low-level preparations for reconfiguration to Constellation 3 contingent upon more information as time progresses.

Decision for Time Period 1: transition to Architecture 3 and begin low-level preparations for reconfiguration to Constellation 3.

4.3 Results: Plan and Pattern

Plan and pattern require evaluating the position according to the perspective of the decision maker and using this information to decide on the best course of action as necessary, while documenting

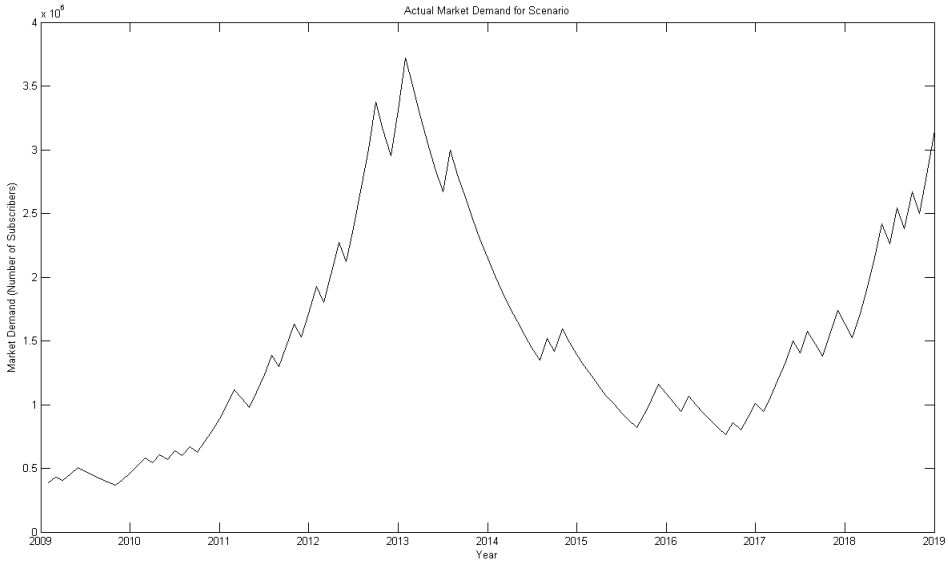


Figure 4.17: Actual Market Demand for the entire system lifetime.

the decision. For the purposes of the case study, plan and pattern will consist of simulating the series of decisions until the system's end of life by evaluating the current information, making a decision, documenting the reasoning and incrementing the time period.

The actual market levels for this example are calculated every month during the 10 years that lead to the horizon. A plot of actual market demand is given in Figure 4.17.

These values are used as inputs into the simulation model to mimic the effects of time in real life.

4.3.1 Strategic Evolution

In this section, the scenario is propagated through time, with a "human in the loop" examining the output maps and making a decision as to the best course of action. This information is fed into the simulation and the next year of "actual" demand levels is propagated. The updated maps are the outputs and the next decision must be made. It is important to note that the horizon window is not truly receding. For the purposes of simplicity and for reducing computation time, the receding horizon is mimicked by updating the horizon energies and re-calculating the terminal entropies as discussed above. This is an assumption that should be relaxed in further studies.

As discussed in the previous section, the immediate transition should be to Architecture 3 with

low-level preparations for reconfiguring to Constellation 3. The outcomes of each time period is discussed below, with documentation on the reasoning for the next decision. The final result will be compared with the results for various other metrics assuming *perfect information* as to the future demand levels.

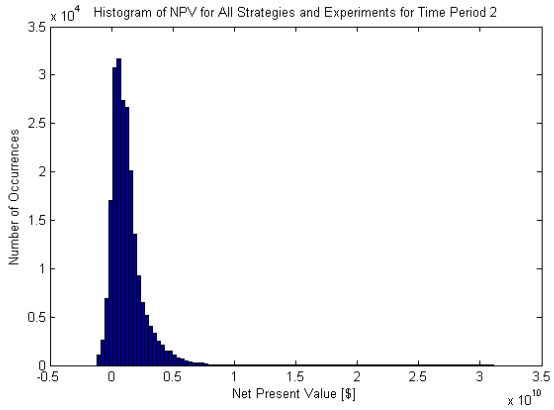
Time Period 2:

The results for Time Period 2 are shown in Figure 4.18.

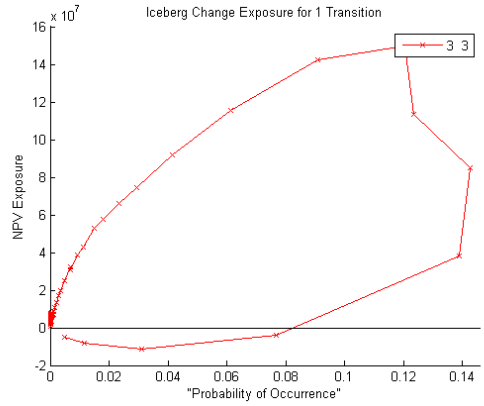
It should be noted that the system is locked into remaining in Architecture 3 due to the first evolution rule. However, it is informative to describe the results of the Strategic Advantage Maps and Iceberg Exposure Graph and to make an updated evaluation on whether to continue preparations for reconfiguring to Constellation 3.

The updated histogram of net present value and the corresponding Iceberg Exposure Graph are shown on the top of Figure 4.18. The Strategic Advantage Maps are shown in the lower graphs, separated by number of transitions considered. The regions are colored by the architectural configuration achieved in the corresponding time period. The immediate transition in this case is Architecture 3 by Evolution Rule 1.

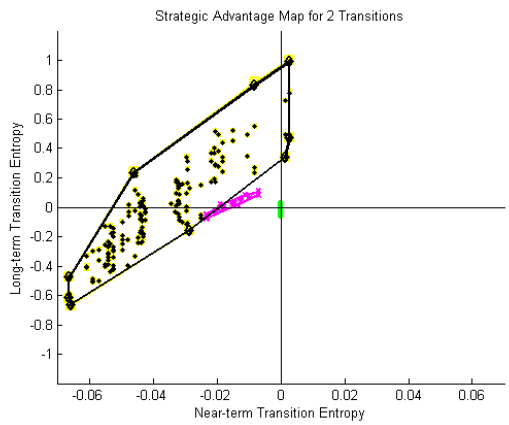
The system has been transitioned to Architecture 3 over the last time period. The old operating expenses were incurred and the benefits gained are based on the previous architecture's usage fees. Since Evolution Rule 1 locks the system into remaining in Architecture 3 for at least one more time period, the first Advantage Map is empty. The second Advantage Map indicates that a slight dominance by Architecture 6 (yellow) in Constellation 3. This isn't a surprise since the only required transition to move from the current system (now Architecture 3) is to reduce the altitude and add satellites. Architecture 7 (black) requires an additional software change. Since the cost of the software change is small relative to the cost of the reconfiguration, it makes sense that any dominance — advantage — would be very small. The other thing to note is that as the view is expanded to more transition periods, there is a noticeable contraction of the instances occurring in Quadrant III. These instances appear to be migrating into Quadrant I and Quadrant II. The trend in contraction and migration would indicate a likely strategic advantage to migrate to Constellation 3. There appears to be some risk involved but the Iceberg Exposure indicates that this risk is minimal. There appears to be far more to gain by continuing preparations to reconfigure to Constellation 3 (likely to Architecture 6) than there is to remain in Architecture 3 or to move to any other architecture.



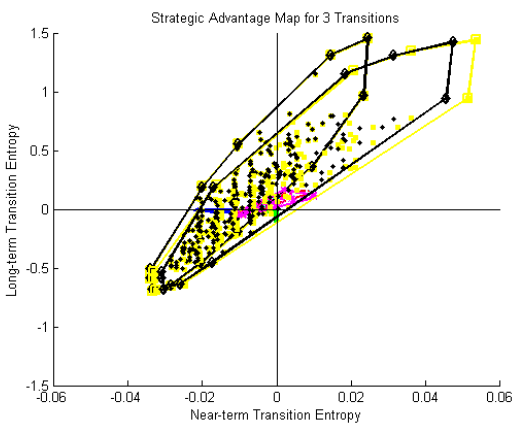
(a) Histogram of Net Present Value



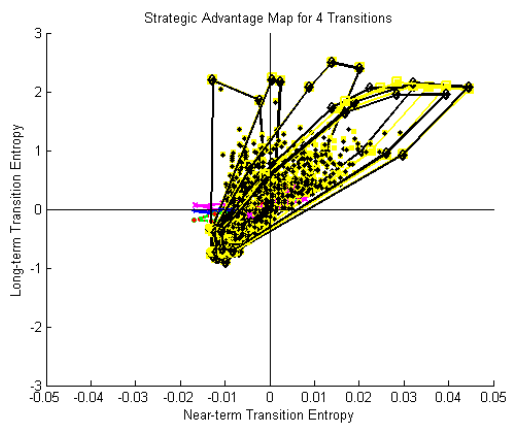
(b) Iceberg Exposure



(c) Two Transitions



(d) Three Transition



(e) Four Transitions

Figure 4.18: The (a) distribution of projected Net Present Value for Time Period 2 and (b) the corresponding Iceberg Exposure looking out one transition. The Strategic Advantage Maps for Time Period 2 looking out (c) two transitions, (d) three transitions, and (e) four transitions.

Thus, by Evolution Rule 1 and by evaluation of updated data, continue in Architecture 3 and step up preparations to reconfigure to Constellation 3 (likely Architecture 6 (yellow)).

Decision for Time Period 2: Continue in Architecture 3 and step up preparations to reconfigure to Constellation 3.

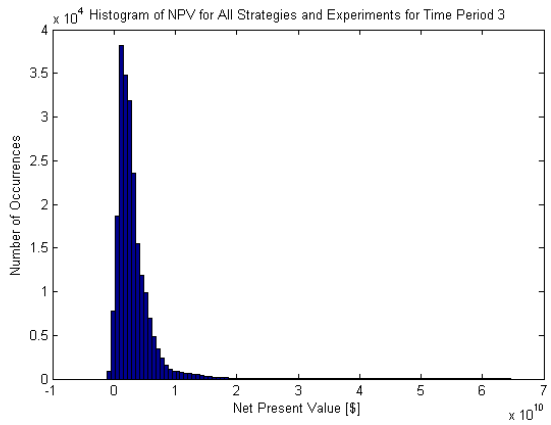
Time Period 3:

The results for Time Period 3 are shown in Figure 4.19.

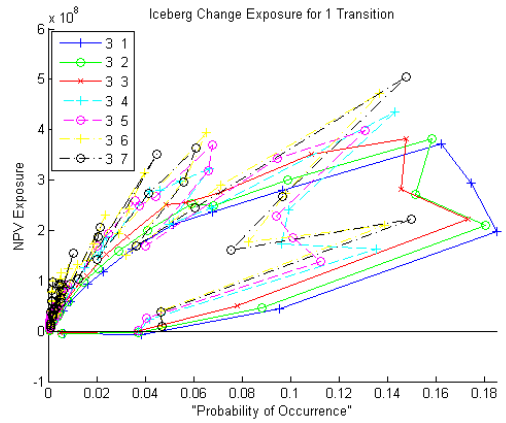
The updated histogram of net present value and corresponding Iceberg Exposure Graph are shown on the top of Figure 4.19 on page 225. The Strategic Advantage Maps are shown in the lower graphs, separated by the number of transitions considered. The regions are colored by the final architecture achieved in the indicated number of transitions.

The first transition map demonstrates what should be expected from this time period. By this point, transitioning to Constellation 3, although it has the near-term disadvantage due to the transition costs, has moved almost completely into Quadrant I, indicating it should always have the long-term advantage from here on out. There is a small region for Constellation 2 architectures (cyan and magenta). Although they do not cost as much to transition, they are strongly dominated along the long-term advantage axis. It is difficult to see, but there are two small striation regions to the left of the origin for transitioning to Architecture 2 (closest to the origin) and to Architecture 1 (further from the origin). This result agrees with expectations.

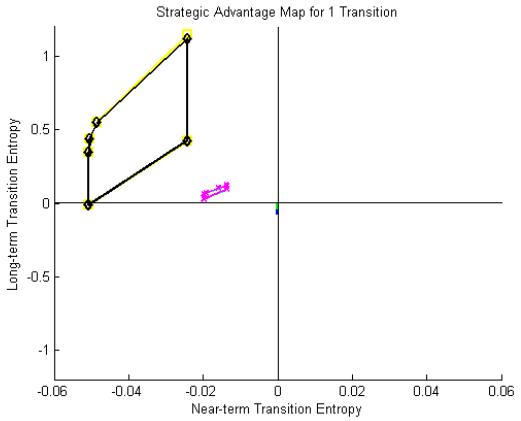
Looking out two transitions, it can be seen that Architecture 6 (yellow) is strongly dominating the other architectures. Figure 4.20 on page 226 provides the Strategic Advantage Maps looking out two transitions and filtered by the immediate transition. The region is slightly more than half into Quadrant II, with the remainder in Quadrant II. Nearly the same thing is true for Architecture 7 (black), but the extent of the dominance (due to usage fees, expense) is becoming more evident. It should also be clear that remaining in Architecture 3 for another time period is also dominated by reconfiguring to Constellation 3. The maximum and minimum long-term positions for all of these are roughly the same, but the Constellation 3 instances do much better along the near-term axis. Again, it is evident the Constellation 2 instances are dominant.



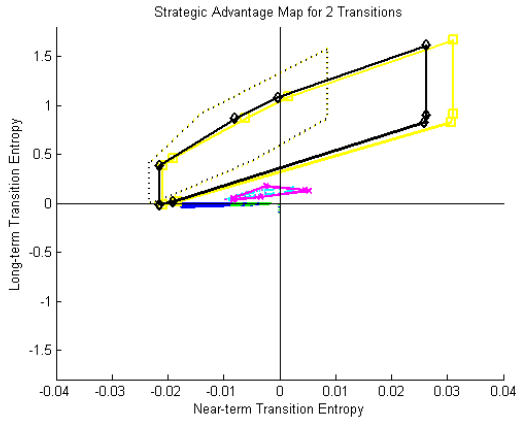
(a) Histogram of Net Present Value



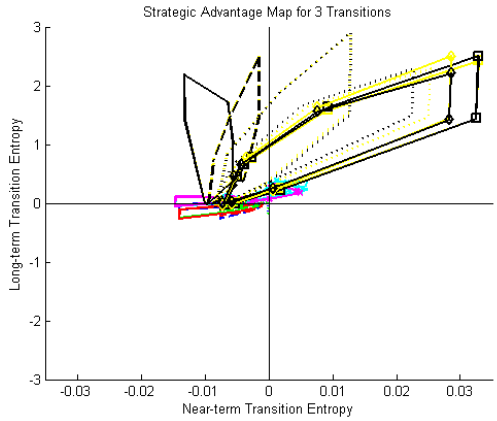
(b) Iceberg Exposure



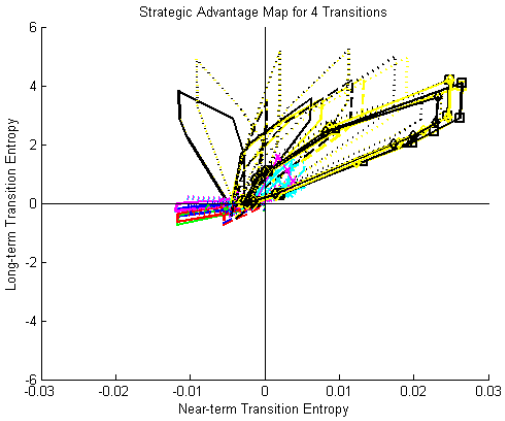
(c) One Transition



(d) Two Transitions

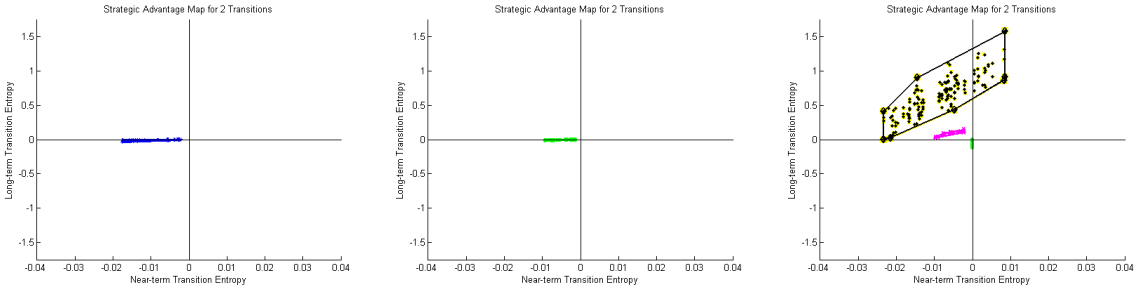


(e) Three Transition



(f) Four Transitions

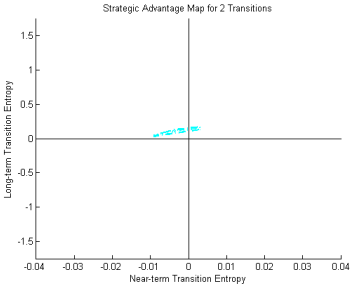
Figure 4.19: The (a) distribution of projected Net Present Value for Time Period 3 and (b) the corresponding Iceberg Exposure looking out one transition. The Strategic Advantage Maps for Time Period 3 looking out (c) one transition, (d) two transitions, (e) three transitions, and (f) four transitions.



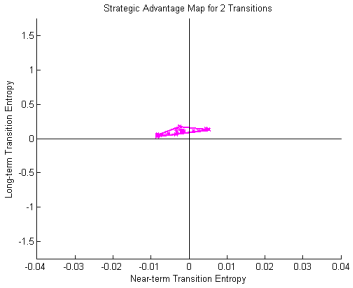
(a) Architecture 1

(b) Architecture 2

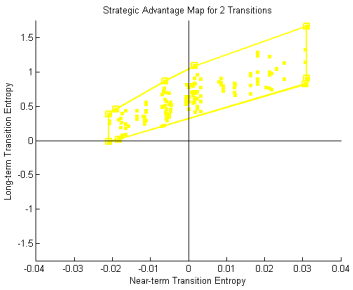
(c) Architecture 3



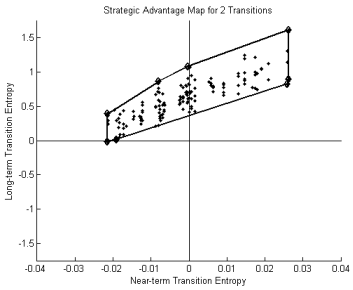
(d) Architecture 4



(e) Architecture 5

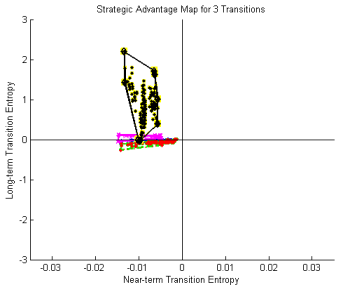


(f) Architecture 6

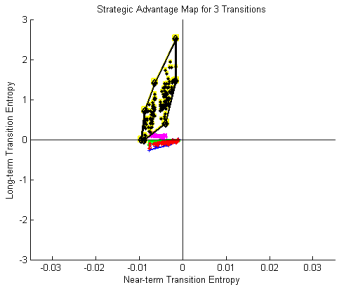


(g) Architecture 7

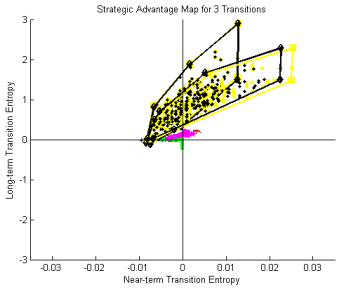
Figure 4.20: The Strategic Advantage Maps for Time Period 3 looking out two transitions and filtered by whether the immediate transition is to (a) Architecture 1, (b) Architecture 2 or (c) Architecture 3 (legacy) in the first constellation, (d) Architecture 4 or (e) Architecture 5 in the second constellation, and (f) Architecture 6 or (g) Architecture 7 in the third constellation. The graphs are colored by the architecture reached in two transitions.



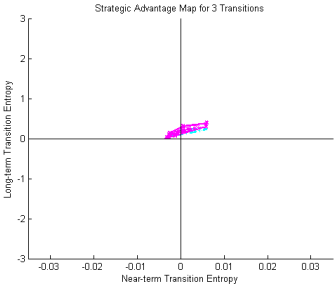
(a) Architecture 1



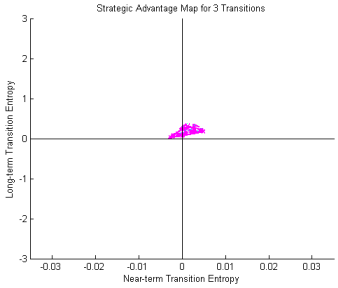
(b) Architecture 2



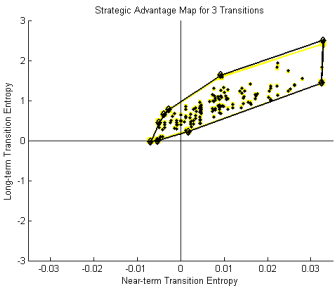
(c) Architecture 3



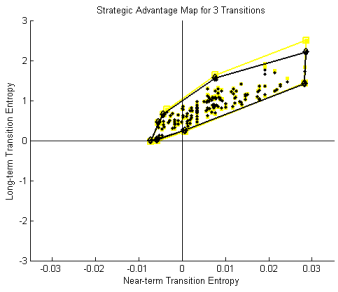
(d) Architecture 4



(e) Architecture 5

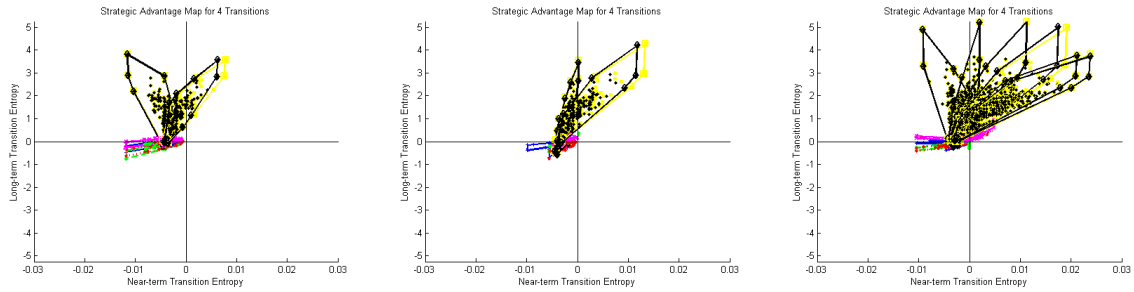


(f) Architecture 6

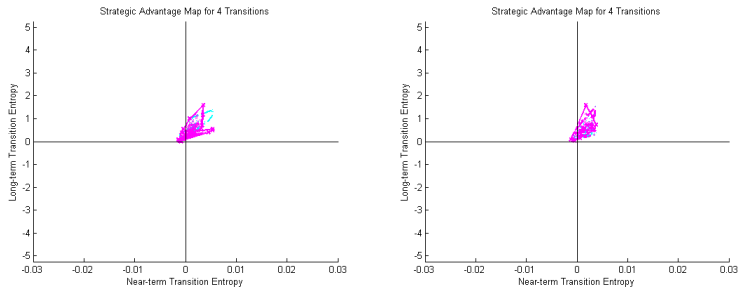


(g) Architecture 7

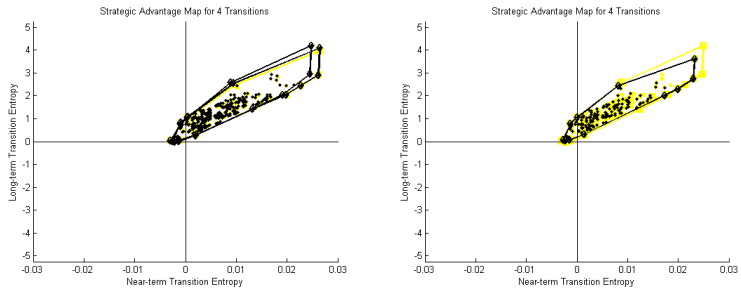
Figure 4.21: The Strategic Advantage Maps for Time Period 3 looking out three transitions and filtered by whether the immediate transition is to (a) Architecture 1, (b) Architecture 2 or (c) Architecture 3 in the first constellation, (d) Architecture 4 or (e) Architecture 5 in the second constellation, and (f) Architecture 6 or (g) Architecture 7 in the third constellation. The graphs are colored by the architecture reached in four transitions.



(a) Architecture 1 (b) Architecture 2 (c) Architecture 3



(d) Architecture 4 (e) Architecture 5



(f) Architecture 6 (g) Architecture 7

Figure 4.22: The Strategic Advantage Maps for Time Period 3 looking out four transitions and filtered by whether the immediate transition is to (a) Architecture 1, (b) Architecture 2 or (c) Architecture 3 in the first constellation, (d) Architecture 4 or (e) Architecture 5 in the second constellation, and (f) Architecture 6 or (g) Architecture 7 in the third constellation. The graphs are colored by the architecture reached in four transitions.

The 3-transition map continues the trends observed for the 2-transition map. The effect of waiting to transition to Constellation 3 is becoming evident. Figure 4.21 on page 227 provides the Strategic Advantage Maps looking out three transitions and filtered by the immediate transition provides insight into the apparent cause of the flower-petal pattern: each petal is the result of a different immediate transition. The achieved architecture in three transitions is separated by color.

The effect of waiting to transition as mentioned in the previous paragraph are even more noticeable

in the 4-transition map (see Figure 4.22) on page 228. Still, it should be evident that an immediate transition to Constellation 6 is the best choice.

Decision for Time Period 3: Choose Architecture 6.

Time Period 4:

The results for Time Period 4 are shown in Figure 4.23 on page 230. Note the "spiky" Iceberg Exposure Graph. The spikes are a result of a "rough" histogram of NPV. The histograms should remain smooth so long as the horizon time is receding. This case study has fixed the horizon time to simplify the simulation and thus the histogram will lose its smoothness the closer the simulation gets to the horizon time.

Again, the decision maker is constrained to stay in this architecture for another period due to Evolution Rule 1, which is reflected in the first transition map. Note that by this point the Iceberg Exposure indicates that all or nearly all of the possibilities end with the system making money based on the NPV. This does NOT mean that the system is guaranteed to break even (there are always random events that have not been accounted for that can pop up and kill the system). Looking forward indicates that for the foreseeable future, the current system (Architecture 6) has the strategic advantage. Naturally, the map no longer accounts for architectures in Constellation 1 or Constellation 2 since they are no longer accessible. However, there could be options added at this point to supplement the option space (see identifying and creating options during system operation), some of which might have the strategic advantage over Architecture 6. For now, however, it is looking pretty good.

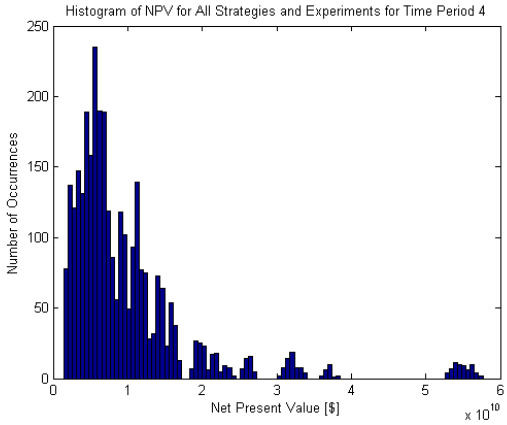
Decision for Time Period 4: Continue with Architecture 6 due to Evolution Rule 1.

Time Period 5:

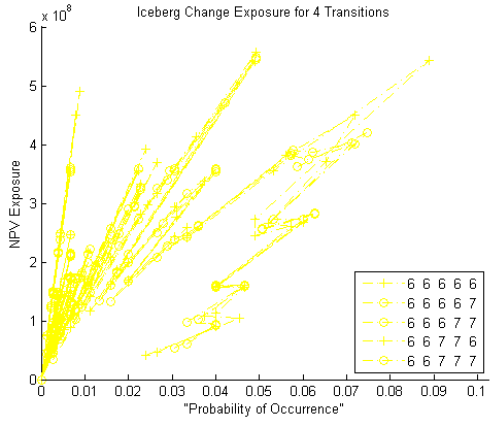
The results for Time Period 5 are shown in Figure 4.24 on page 231. Figure 4.25 on page 232 gives the Strategic Advantage Maps filtered by the number of transitions considered (row) and the immediate transition (column).

Here, there is not much to say. The current legacy architecture clearly dominates the option space. Unless additional options are identified and/or created, the best next transition should be evident.

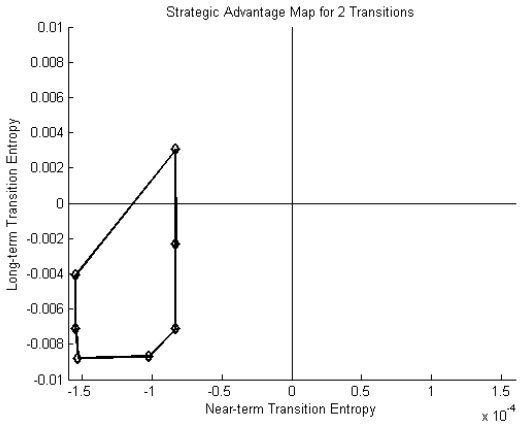
Decision for Time Period 5: Continue with Architecture 6.



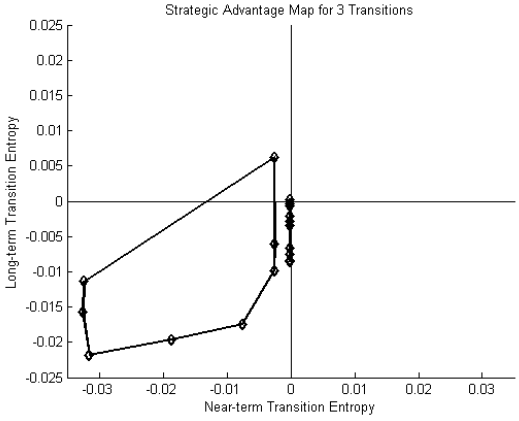
(a) Net Present Value



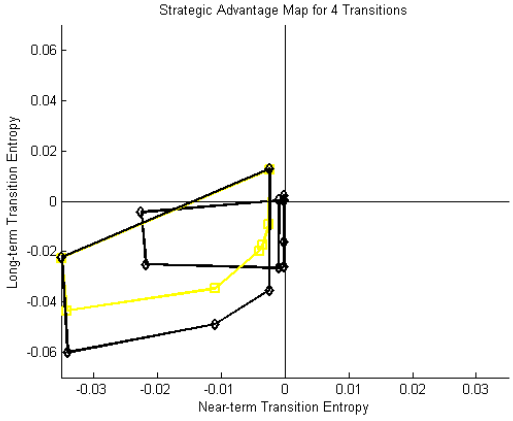
(b) Iceberg Exposure



(c) Two Transitions

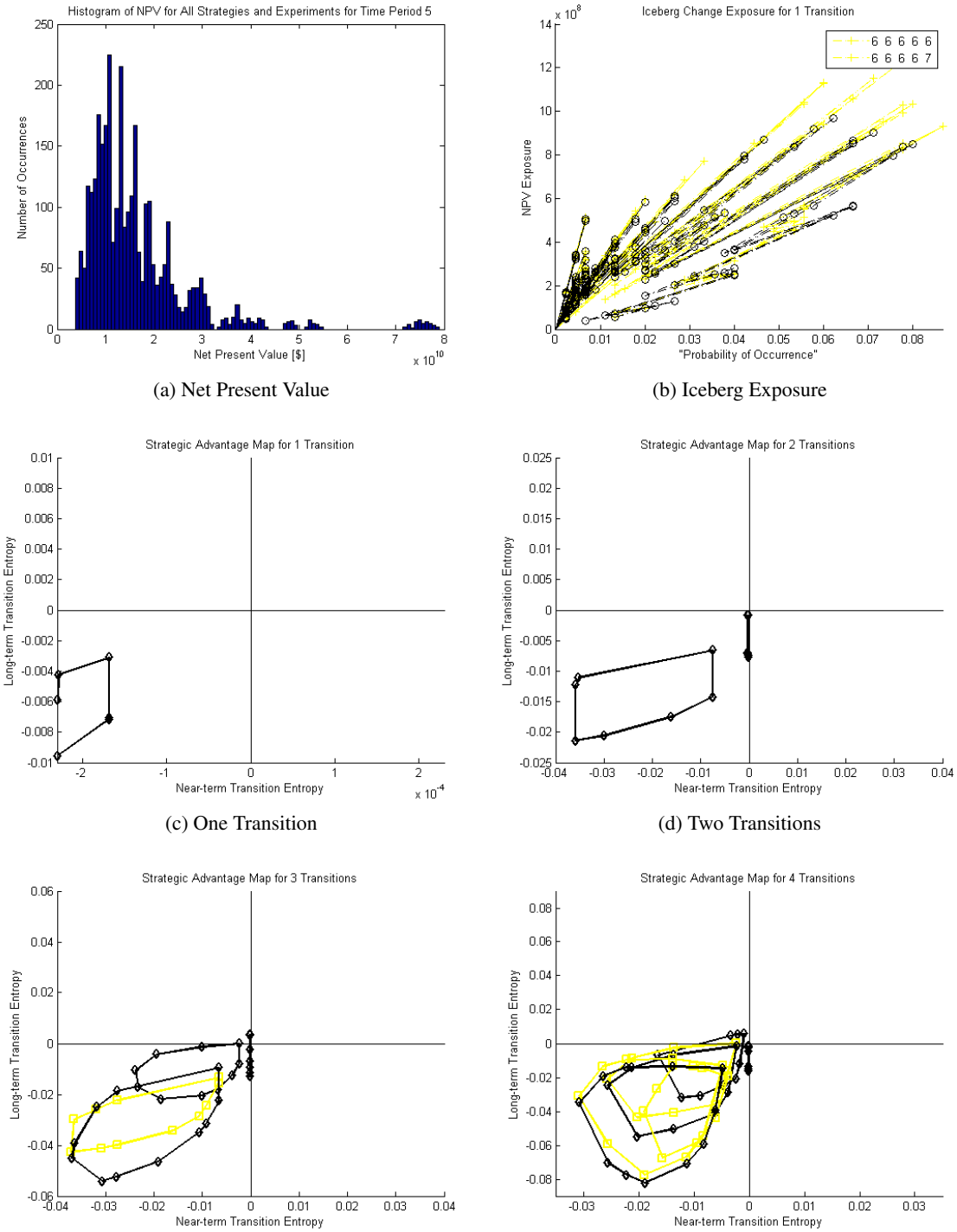


(d) Three Transitions



(e) Four Transitions

Figure 4.23: The (a) distribution of projected Net Present Value for Time Period 4 and (b) the corresponding Iceberg Exposure looking out four transitions. The Strategic Advantage Maps for Time Period 4 looking out (c) two transitions, (d) three transitions, and (e) four transitions.



(a) Net Present Value

(b) Iceberg Exposure

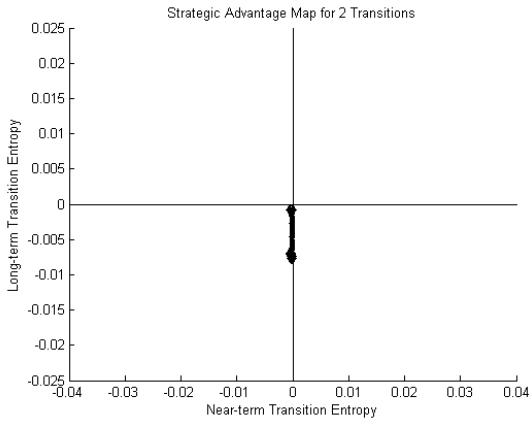
(c) One Transition

(d) Two Transitions

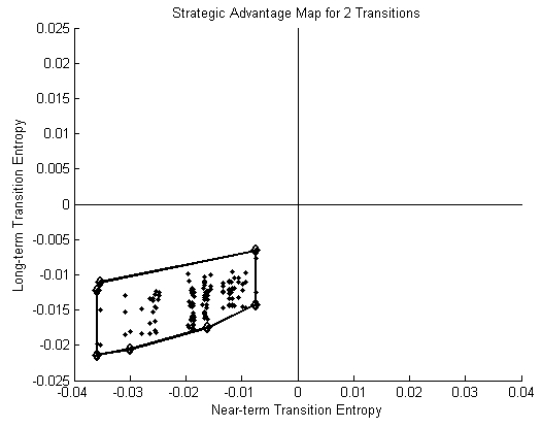
(e) Three Transitions

(f) Four Transitions

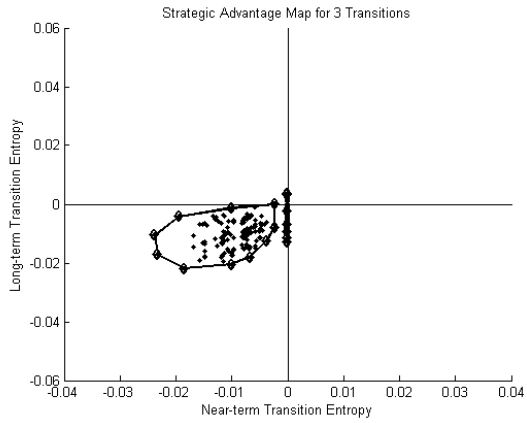
Figure 4.24: The (a) distribution of projected Net Present Value for Time Period 5 and (b) the corresponding Iceberg Exposure looking out four transitions. The Strategic Advantage Maps for Time Period 5 looking out (c) one transition, (d) two transitions, (e) three transitions, and (f) four transitions.



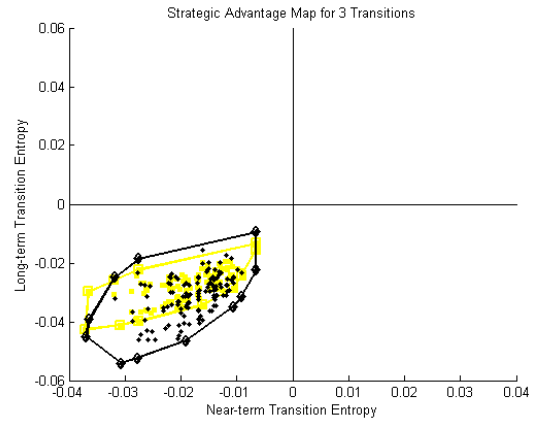
(a) Two Transitions, Architecture 6



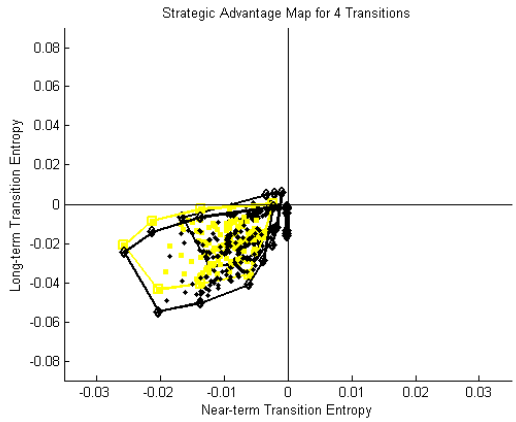
(b) Two Transitions, Architecture 7



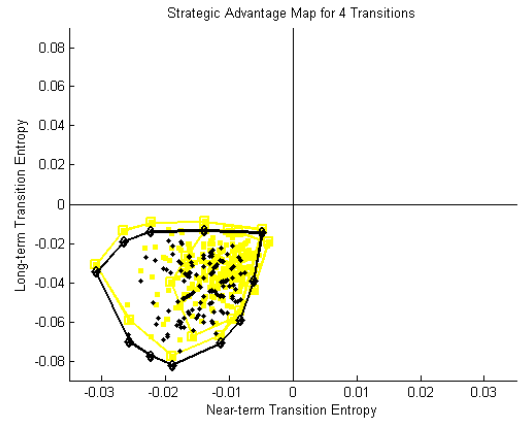
(c) Three Transitions, Architecture 6



(d) Three Transitions, Architecture 7



(e) Four Transitions, Architecture 6



(f) Four Transitions, Architecture 7

Figure 4.25: Looking out two transitions in Time Period 5, the effect of transitioning to (a) Architecture 6 and (b) Architecture 7. Looking out three transitions in Time Period 5, the effect of transitioning to (c) Architecture 6 and (d) Architecture 7. Looking out four transitions in Time Period 5, the effect of transitioning to (e) Architecture 6 and (f) Architecture 7.

Time Period 6:

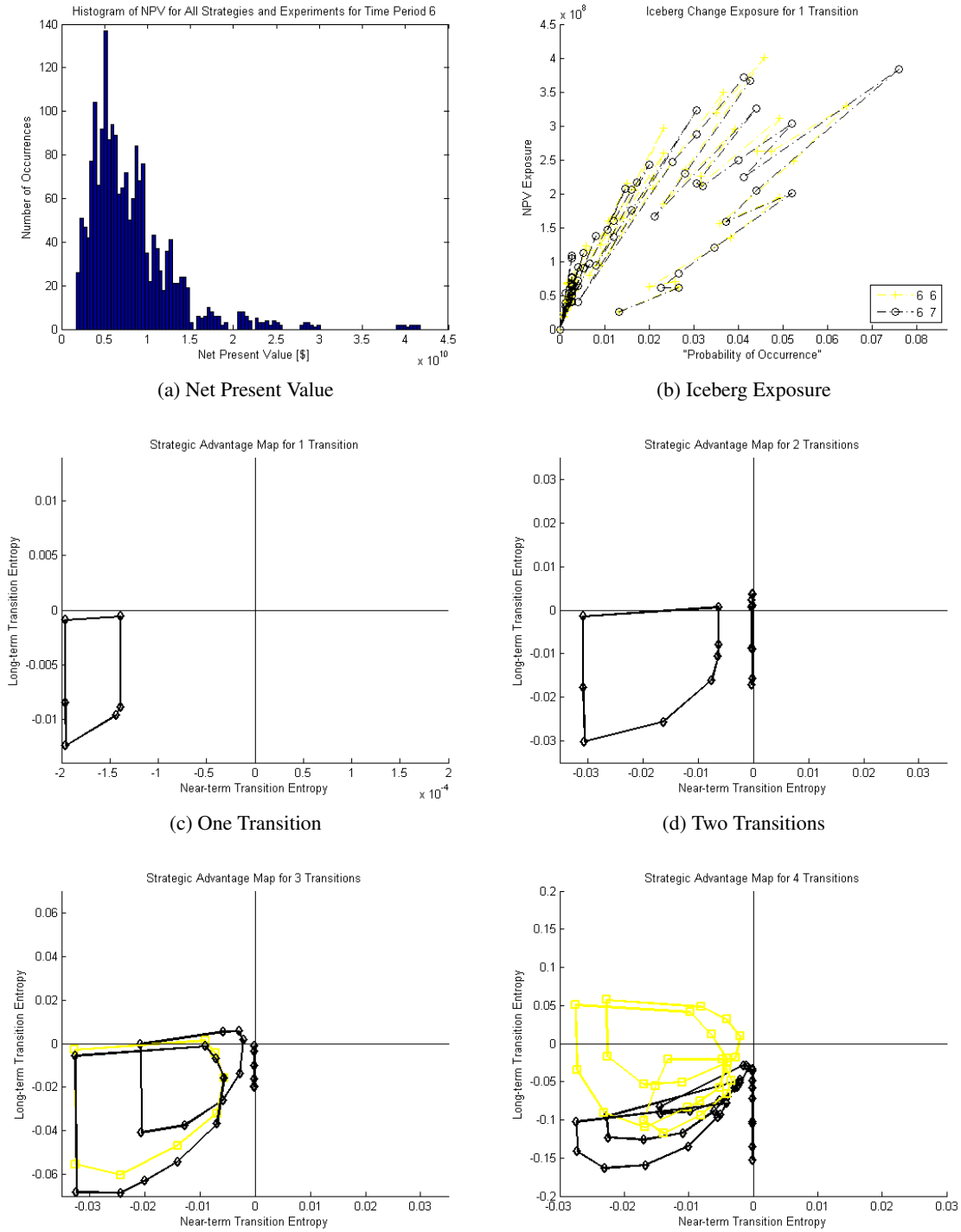
The results for Time Period 6 are shown in Figure 4.26 on page 234. Figure 4.27 on page 235 gives the Strategic Advantage Maps filtered by the number of transitions considered (row) and the immediate transition (column).

There is a creeping trend beginning that indicates that there might be a more optimal transition path (compare Figure 4.24 (d) and Figure 4.26 (d)). There is an indication that briefly transitioning to Architecture 7 (black) might be the optimal choice, but at this point, the possibility is weak (still strongly dominated by the legacy system). Looking at Figure 4.17 on page 221, assuming that only the period 2009 to 2015 is known, there is a definite decline in the number of subscribers. If this level drops enough, then Architecture 7 will move from Business Area 11 to Business Area 7 at which point the current usage fee of $75 will jump to $135. However, this will not happen unless the number of subscribers drops to below 200,000. What this map is saying is that the rate of decline is significant enough that there is a small possibility that the system will need to leverage the option to switch business areas, but that the time has not yet come. If this were to happen and the number of subscribers rose to the point that the value of this switch is threatened, the decision makers will need to investigate whether there is another accessible architecture that can provide this new service level (likely will need to be reconfigure to a lower altitude).

Decision for Time Period 6: Continue with Architecture 6, start looking for new options at a low-level.

Time Period 7:

The results for Time Period 7 are shown in Figure 4.28 on page 236. Figure 4.29 on page 237 gives the Strategic Advantage Maps filtered by the number of transitions considered (row) and the immediate transition (column).



(a) Net Present Value

(b) Iceberg Exposure

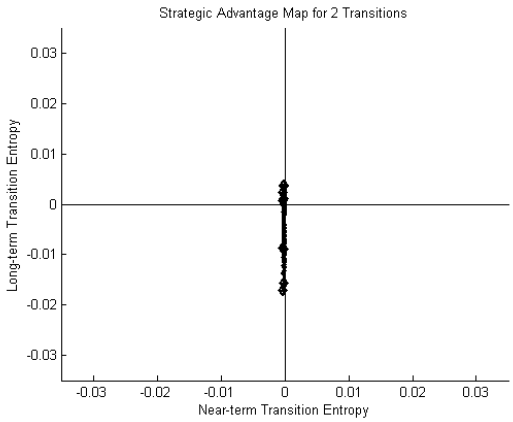
(c) One Transition

(d) Two Transitions

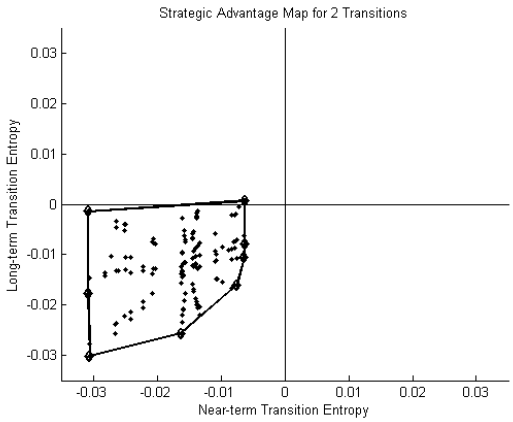
(e) Three Transitions

(f) Four Transitions

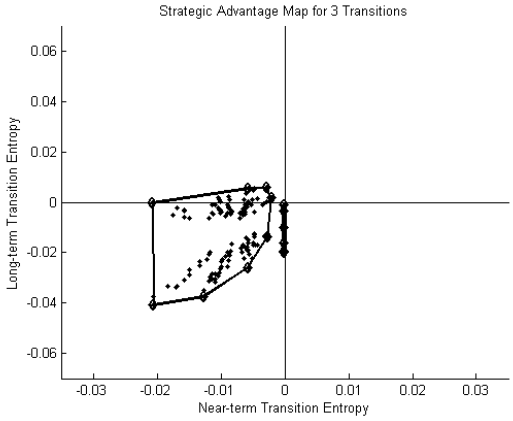
Figure 4.26: The (a) distribution of projected Net Present Value for Time Period 6 and (b) the corresponding Iceberg Exposure looking out four transitions. The Strategic Advantage Maps for Time Period 6 looking out (c) one transition, (d) two transitions, (e) three transitions, and (f) four transitions.



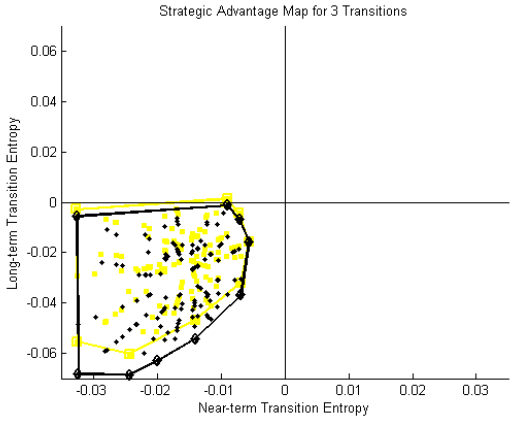
(a) Two Transitions, Architecture 6



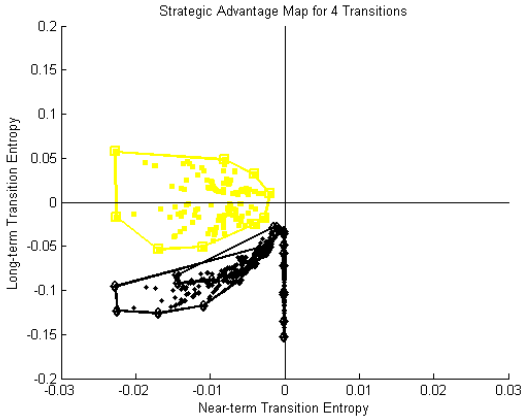
(b) Two Transitions, Architecture 7



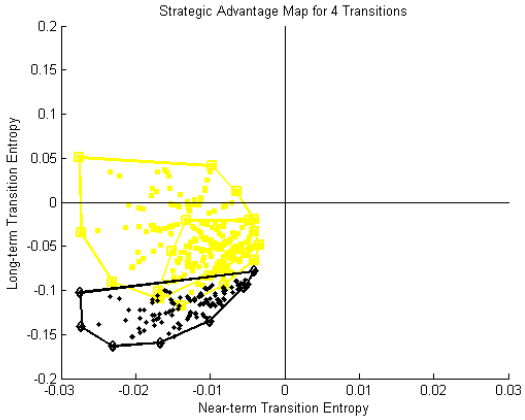
(c) Three Transitions, Architecture 6



(d) Three Transitions, Architecture 7

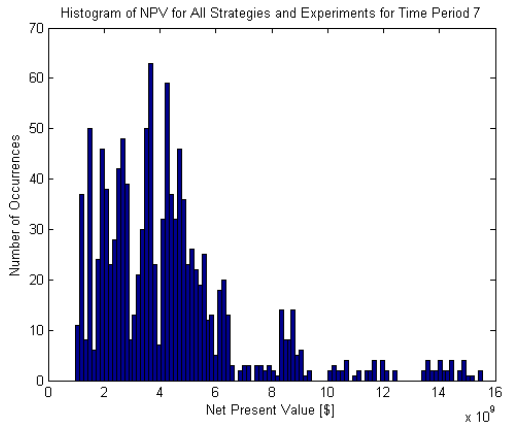


(e) Four Transitions, Architecture 6

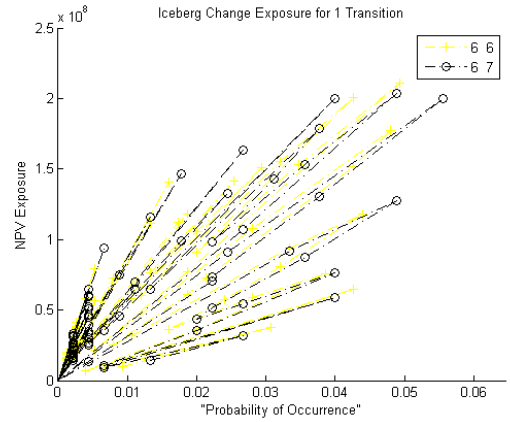


(f) Four Transitions, Architecture 7

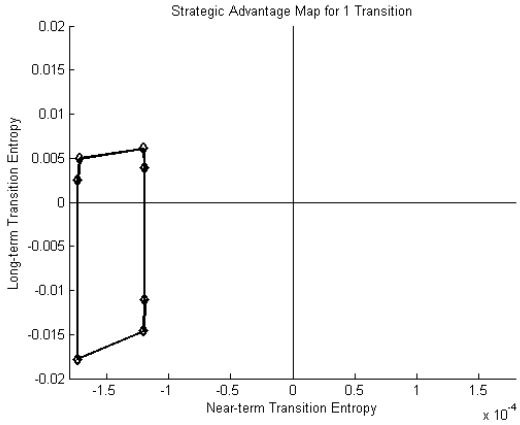
Figure 4.27: Looking out two transitions in Time Period 6, the effect of transitioning to (a) Architecture 6 and (b) Architecture 7. Looking out three transitions in Time Period 6, the effect of transitioning to (c) Architecture 6 and (d) Architecture 7. Looking out four transitions in Time Period 6, the effect of transitioning to (e) Architecture 6 and (f) Architecture 7.



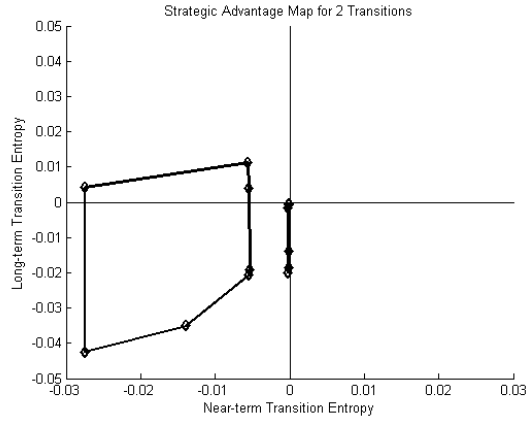
(a) Net Present Value



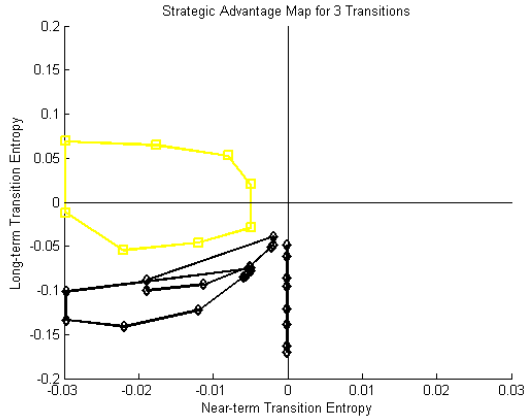
(b) Iceberg Exposure



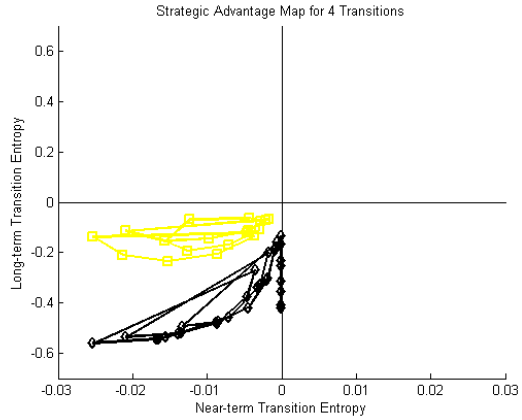
(c) One Transition



(d) Two Transitions

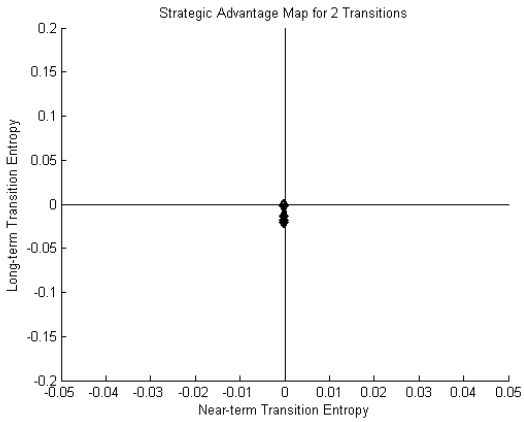


(e) Three Transitions

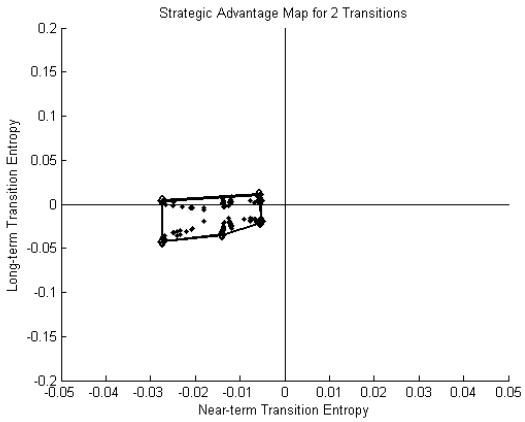


(f) Four Transitions

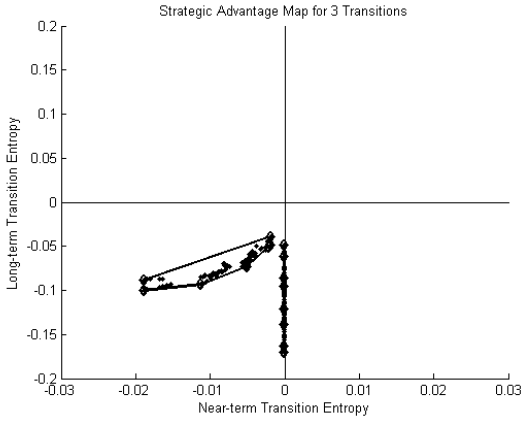
Figure 4.28: The (a) distribution of projected Net Present Value for Time Period 7 and (b) the corresponding Iceberg Exposure looking out four transitions. The Strategic Advantage Maps for Time Period 7 looking out (c) one transition, (d) two transitions, (e) three transitions, and (f) four transitions.



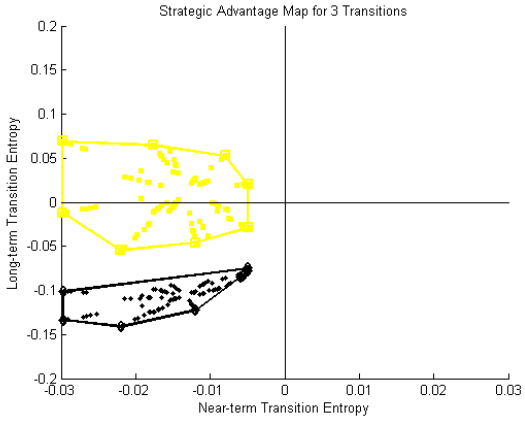
(a) Two Transitions, Architecture 6



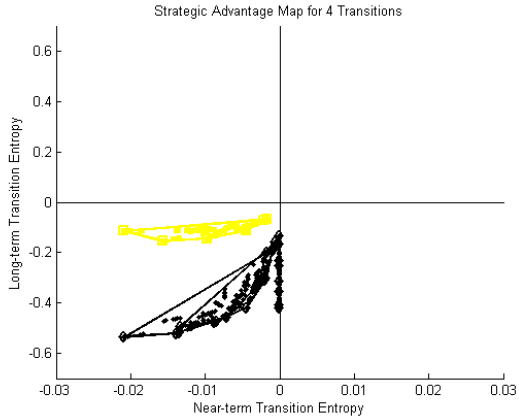
(b) Two Transitions, Architecture 7



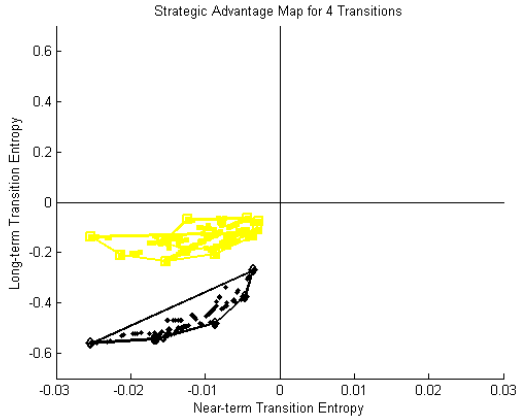
(c) Three Transitions, Architecture 6



(d) Three Transitions, Architecture 7



(e) Four Transitions, Architecture 6



(f) Four Transitions, Architecture 7

Figure 4.29: Looking out two transitions in Time Period 7, the effect of transitioning to (a) Architecture 6 and (b) Architecture 7. Looking out three transitions in Time Period 7, the effect of transitioning to (c) Architecture 6 and (d) Architecture 7. Looking out four transitions in Time Period 7, the effect of transitioning to (e) Architecture 6 and (f) Architecture 7.

Here it is clear that the number of subscribers has dropped even more. The map is showing an even stronger likelihood of needing to switch to Architecture 7. It is also clear from looking at the one-transition and two-transition maps that this time is not now. Fortunately, software changes can be made relatively quickly (by the end of the next transition period is the assumption), so there is no need to act just yet. The possibility is looming a few periods down the line, but there is also a likelihood the trend will reverse itself and expending the extra resources to move to an architecture that only performs better for a significantly lower demand level is not a wise move. Still, the map indicates to keep an eye out and to possibly start looking at other options.

Decision for Time Period 7: Continue with Architecture 6, but increase the resources looking into alternatives.

Time Period 8:

The results for Time Period 8 are shown in Figure 4.30 on page 239. Figure 4.31 on page 240 gives the Strategic Advantage Maps filtered by the number of transitions considered (row) and the immediate transition (column).

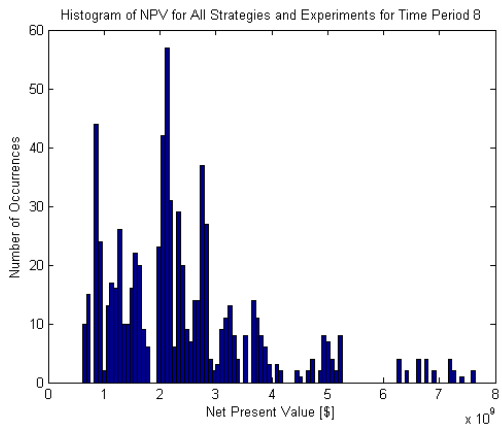
In these maps, it is evident that the decline in subscriber demand has stabilized, and the likelihood of needing to switch architectures has decreased. Now, here is where having a fully receding horizon is important. It is unlikely that the map would have relaxed this much given the trend in the actual market data. This is almost certainly due to being so close to the horizon and the relatively fixed calculations for the terminal entropy. However, the initial transition graph gives the most clue and it indicates that the likelihood is still there but has been reduced.

Based on this information, the decision is to continue with Architecture 6, and reduce the resources looking into alternatives. The system isn't out of the woods yet, but there is no need to expend much more.

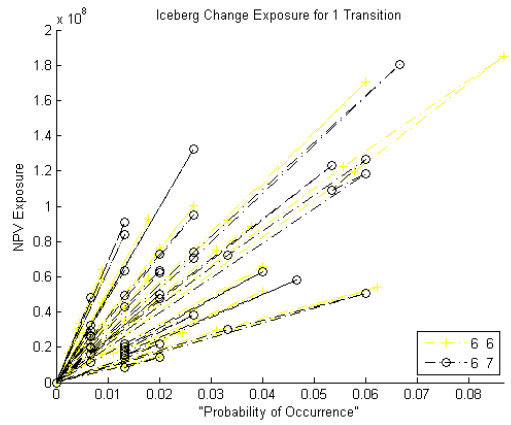
Decision for Time Period 8: Continue with Architecture 6, reduce resources looking into alternatives.

Time Period 9:

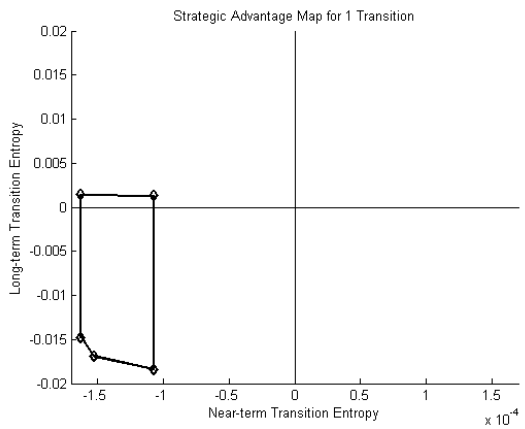
The results for Time Period 9 are shown in Figure 4.32 on page 241. Figure 4.33 on page 241 gives the Strategic Advantage Maps filtered by the immediate transition.



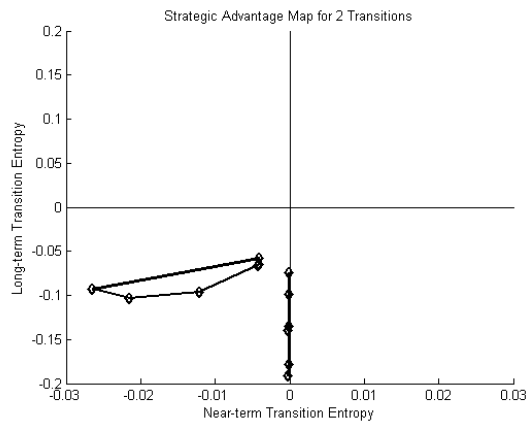
(a) Net Present Value



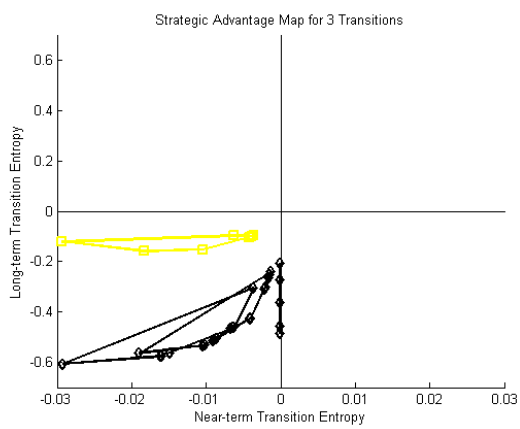
(b) Iceberg Exposure



(c) One Transition

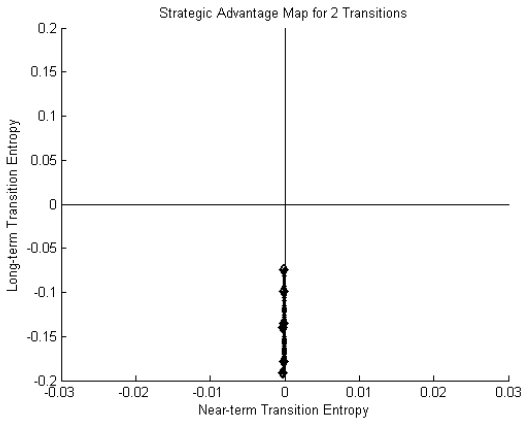


(d) Two Transitions

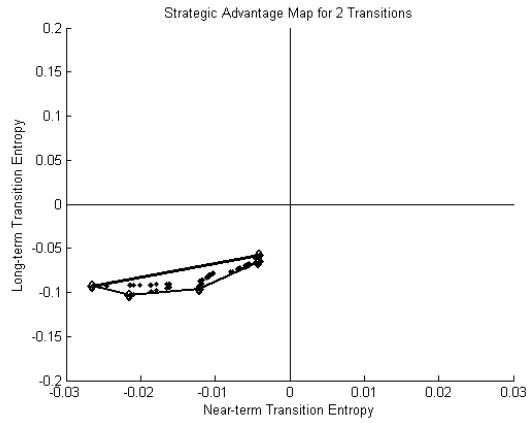


(e) Three Transitions

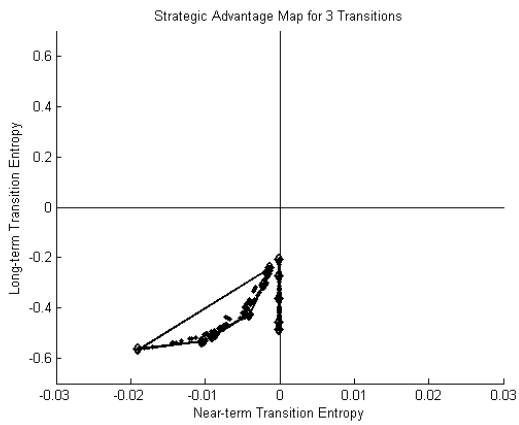
Figure 4.30: The (a) distribution of projected Net Present Value for Time Period 8 and (b) the corresponding Iceberg Exposure looking out one transition. The Strategic Advantage Maps for Time Period 8 looking out (c) one transition, (d) two transitions, and (e) three transitions.



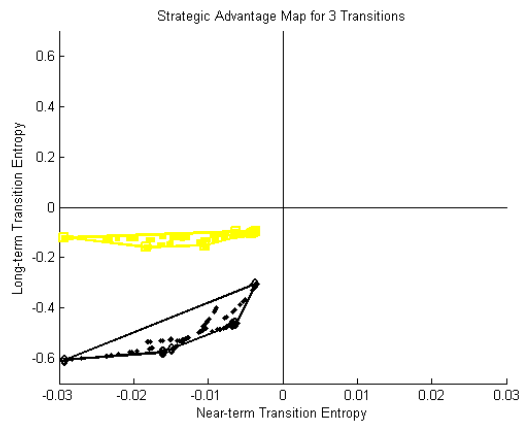
(a) Two Transitions, Architecture 6



(b) Two Transitions, Architecture 7



(c) Three Transitions, Architecture 6



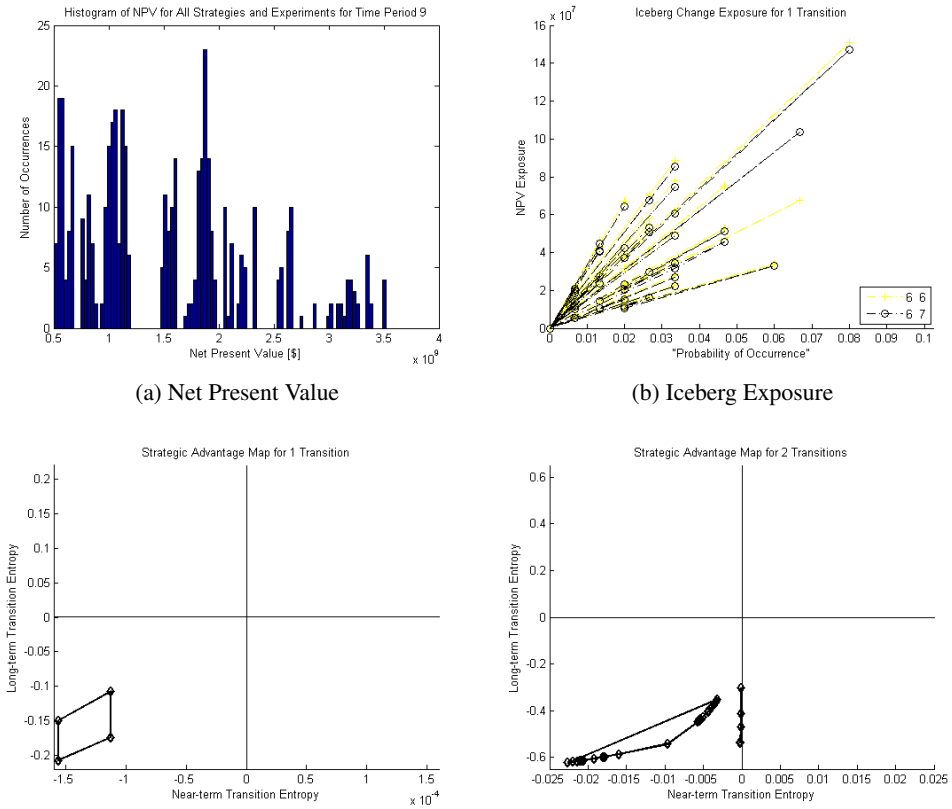
(d) Three Transitions, Architecture 7

Figure 4.31: Looking out two transitions in Time Period 8, the effect of transitioning to (a) Architecture 6 and (b) Architecture 7. Looking out three transitions in Time Period 8, the effect of transitioning to (c) Architecture 6 and (d) Architecture 7.

Here it is quite clear that the dropping subscriber trend has reversed itself. However, the impact of the terminal entropies and lack of a truly receding horizon probably have a lot to do with this. A look at the market data supports the conclusion that the trend has largely reversed itself and it is probably safe to back off of funding for alternatives. But to know for sure, the model would need to incorporate a receding horizon.

The Strategic Advantage maps suggest decision makers should continue with Architecture 6 until the horizon. Since the horizon is not fully modeled, it is impossible to say what the horizon decision will be, but it is likely to be that the system should remain in Architecture 6.

Decision for Time Period 9: Continue with Architecture 6 until the horizon.



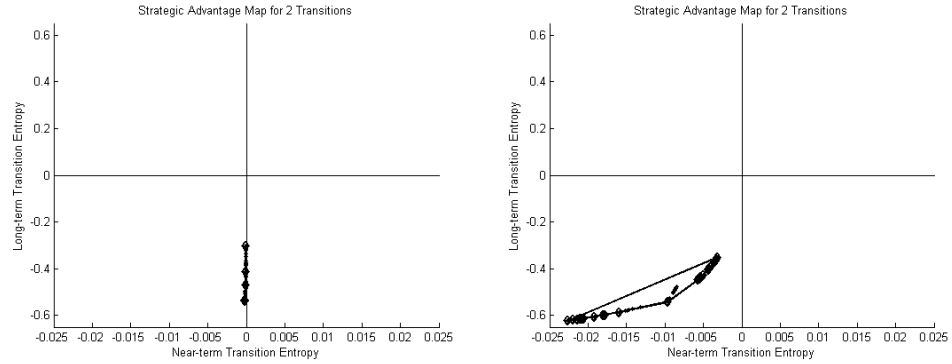
(a) Net Present Value

(b) Iceberg Exposure

(c) One Transition

(d) Two Transitions

Figure 4.32: The (a) distribution of projected Net Present Value for Time Period 9 and (b) the corresponding Iceberg Exposure looking out one transition. The Strategic Advantage Maps for Time Period 9 looking out (c) one transition, and (d) two transitions.



(a) Architecture 6

(b) Architecture 7

Figure 4.33: Looking out two transitions in Time Period 9, the effect of transitioning to (a) Architecture 6 and (b) Architecture 7.

Effect of Perfect Information

Method	Path	Net Present Value	Life Cycle Cost	Internal Rate of Return
Maximum Benefit	1-3-3-6-6-6-6-6-6-6-6	$4.1914 billion	$2.8930 billion	37.37%
Minimum Life Cycle Cost	1-1-1-1-1-1-1-1-1-1-1	$9.8677 million	$2.1229 billion	10.1%
Maximum Net Present Value	1-3-3-6-6-6-6-6-6-6-6	$4.1914 billion	$2.8930 billion	37.37%
Benefit-Cost	1-3-3-3-3-3-3-3-3-3-3	$3.4210 billion	$2.1236 billion	36.6%
Strategic Evolution	1-3-3-6-6-6-6-6-6-6-6	$4.1914 billion	$2.8930 billion	37.37%

Table 4.16: The effect of perfect information on the optimal transition path for various metrics and compared with the outcome of the Strategic Evolution of Systems method for the demand scenario considered in this case study.

Thus, the optimal path chosen by the Strategic Evolution of Systems is: 1-3-3-6-6-6-6-6-6-6-6.

4.3.2 Effect of Perfect Information

This section briefly describes the effect of perfect information on the optimal path for different desirability metrics and compares the results with the outcome of the Strategic Evolution of Systems simulation.

What is the optimal path *given perfect information*?

If it were possible to look into a crystal ball in 2009 and to know exactly what the subscriber demand would be for the life of the system, then it would be possible to know the optimal path the system should take through time. The "optimal" path depends on the metric under consideration. This section highlights the results of propagating the actual demand scenario through time, accumulating the revenues and expenses and evaluating the outcomes using various established metrics. The results are summarized in Table 4.16

Maximum Benefit

If the only concern is to maximize the present value of the benefit as seen in 2009, then the optimal path is 1-3-3-6-6-6-6-6-6-6-6. In this case, the Net Present Value is $4.1914 billion. The present value of the life cycle cost is $2.8930 billion. The internal rate of return is 37.37%.

Minimum Life Cycle Cost

The minimum life cycle cost (in constant dollars) occurs with path number 1, with a value of $2.1229 billion. This path is: 1-1-1-1-1-1-1-1-1-1-1, the maintain "as is" legacy path. Since the actual

demand never saturates the legacy system to the point it can no longer achieve a business area, it is valid to consider this architecture path. Interestingly, the corresponding Net Present Value for this path is $9.8677 million, which indicates that the system achieves a return on the investment. This translates to an internal rate of return of approximately 10.1%.

Maximum Net Present Value

The optimal path based on maximizing the Net Present Value is also 1-3-3-6-6-6-6-6-6-6-6. Again, the Net Present Value is $4.1914 billion with a corresponding internal rate of return of 37.37%.

Benefit-Cost

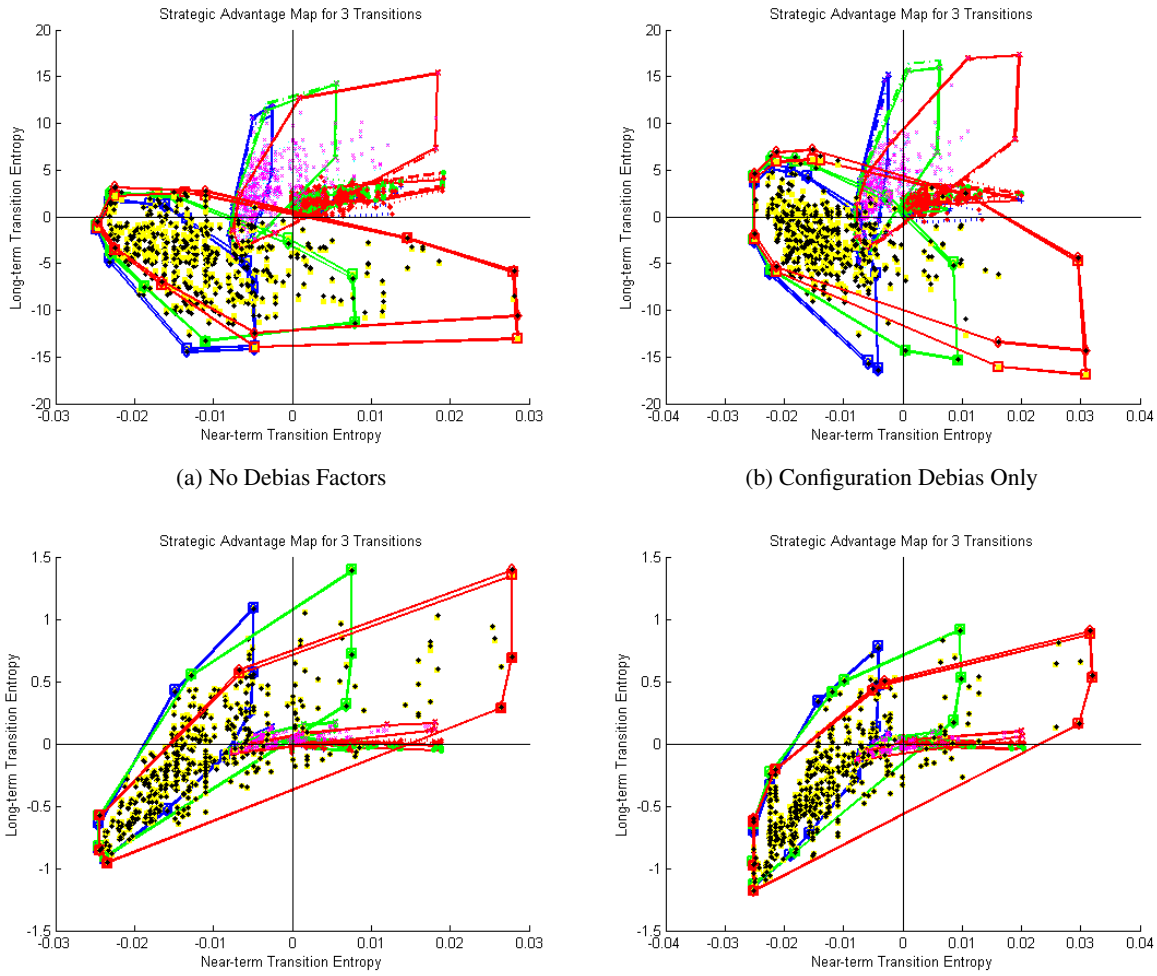
If the metric is benefit-cost, then the optimal path is 1-3-3-3-3-3-3-3-3-3-3, giving a value of 2.6109. The Net Present Value is $3.4210 billion and the life cycle cost is $2.1236 billion. The internal rate of return comes out to roughly 36.6%.

4.3.3 Effect of Debiasing Factors

It has been mentioned that the entropy formulation is sensitive to the choice of debiasing factors. This section briefly demonstrates this statement to be true.

Figure 4.34 shows the effect of the two debiasing factors — for the configuration term and the number of option paths — separately and combined. The example is for the first time period (the initial Strategic Advantage Maps) looking out three transitions. The three architectures comprising Constellation 1 are blue, green, and red; the two architectures comprising Constellation 2 are cyan and magenta; and the two architectures comprising Constellation 3 are yellow and black.

The most noticeable difference is the significant change in shape and scale when the number of option paths is scaled according to the debias factor described in the Weightings section. The long-term entropy axis shrinks from 20 to 1.5. The near-term axis is consistent because of the way the near-term entropy is estimated. Since the number of options available to the lower constellations are artifically scoped — seven options when in Constellation 1 reduces to four options in Constellation 2 and further reduces to two options in Constellation 3 — the number of option paths are also artificially scoped and this can bias the entropy formulation in favor of remaining in Constellation 1 when this may be a poor choice relative to the other constellations.



(a) No Debias Factors

(b) Configuration Debias Only

(c) Number of Option Paths Debias Only

(d) Both Debias Factors

Figure 4.34: A demonstration of the effect of the debias factors used in this case study. The figures represent Time Period 1 looking out three transitions with (a) no debias factors, (b) the configuration debias factor only, (c) the debias factor for the number of option paths only, and (d) both debias factors considered simultaneously.

The configuration term is not nearly as significant though it does appear to contract the graph slightly toward the origin along the long-term axis. Without the option paths debiasing factor, the configuration term separates the regions specificed by the different (immediate transition) architectures in Constellation 1. Combining the option paths debiasing factor with the configuration term debiasing factor appears to level out the long-term entropies along the top of the graph.

4.4 Discussion

In this example, the Strategic Evolution of System methodology, incorporating the Strategic Advantage and Iceberg Exposure maps, indicates the optimal path through the option space. With imperfect information, the visualization tools and position information helped to guide the decision maker toward the best set of choices.

It is not claimed that this methodology will return the optimal path in all cases, but the example demonstrates the usefulness of the tools and the methodology. The primary caveat to this result is that it is completely dependent on the set of options chosen by the decision makers (or support analysts) to be considered in the final analysis. Thus, it is clearly important that the Perspective steps be re-evaluated frequently so as to make sure that the best set of options are in play at all times.

The set of decisions were chosen because of their value in demonstrating the methodology and the visualization tools.

The visualization tools are currently limited to about seven options. Including more would clutter the Strategic Advantage Maps and make it difficult to see the patterns. Future work could investigate better ways of visualizing the position data, particularly if additional sources of uncertainty are modeled. There may also be mathematical relationships within the outcomes that could provide qualitative support to the decision makers. At the moment, the manner of reading the maps and evaluating the meaning is completely qualitative. Although the maps are quite illustrative, it is difficult to understand the meaning of the patterns. This will be an acquired skill, as it requires deep understanding of the nature of the transition rules, the evolution rules and the manner in which the entropy metric is calculated.

The method is similar to decision analysis in the way paths (decisions) are explicitly laid out. The Strategic Evolution of Systems, however, does not require a decision node to be followed by separate chance nodes representing each source of uncertainty and limited to a small number of probabilistic outcomes (this is difficult to model for continuous random variables). The inter-relationships and path dependencies of the option space are modeled and a monte carlo sampling of the uncertainties over time are fed in and incorporated into the energy propagation. The result is something akin to the wave-particle experiment, in which a photon of light is fired at two narrow slits and the distribution of the pattern on a wall behind the slits is measured. The sampling of the uncertainties

is fired into the model of the option space and the distribution of the desirability of its response to the uncertainty over time is captured on the other side.

4.5 Simulation Benchmarking

The core simulation model was benchmarked to Iridium (the famous global low-earth polar constellation with publicly-available information). It is also the most mature MF-TDMA system.[75] The comparisons are summarized in Table 4.17.

The relevant variables were set to known and best-guessed Iridium values.

The costing is consistent over a five-year lifetime. Lloyd Wood published a figure of $3.7 billion for the Iridium life cycle cost over five years[76]. The simulator estimates $3.51 billion between the development, manufacturing and launch costs and the expected operating costs over five years.

The simulation predicts 17 launches with a Long-March 2C rocket. In reality, there were 15 launches, consisting of Proton, Delta 2 and Long March rockets.

The satellite Effective Isotropic Radiated Power (EIRP) is consistent. The simulator comes up with 56 Decibel (dB)/sat compared to 57.04 dB/sat.[77]

There is a discrepancy in the spacecraft mass: the simulator predicts a satellite weighing 537 kilograms while the Iridium satellites were 700 kilograms.

The number of Frequency Division Multiple Access (FDMA) channels were estimated at 120.0298 compared to 120.[75] Similarly, the number of Time Division Multiple Access (TDMA) duplex channels were found to be 4.2875 rather than 4. The number of satellite channels were found to be 1029.3 compared to 1100. The packet length, if the simulation uses the RIP routing protocol and stop-and-wait ARQ, was found to be 490 bits instead of 414 bits.

The simulation predicts that given 60,000 subscribers (the actual subscriber number published by Iridium), with a receiver diameter of 0.1-meters over a network using Routing Information Protocol (minimum hop routing), stop-and-wait ARQ, and distributed routing, then the average Round-trip Delay (RTD) is expected to be 0.1807 seconds, the jitter to be 0.1920 seconds and the loss to be 0.28 percent. This quality of service, according to the simulation, supports bulk data transfers. However, if we relax the 60 millisecond (ms) jitter requirement to encompass the simulated jitter, then minimum telephony and interactive data business areas are included. Jitter effects could account for the

Iridium Benchmark

Description	Iridium Value	Simulation Value	Units
5-year life cycle cost	3.7	3.51	$ billion
Number of launches	15	17	-
Launch vehicle	Proton, Delta 2 and Long March	Long-March 2C	-
Spacecraft mass	700	537	kg
Number of FDMA channels	120	120.0298	-
Number of TDMA duplex channels	4	4.2875	-
Number of satellite channels	1100	1029.3	-
Packet length	414	490	bits
Satellite EIRP	57.04	56	dB
Round-trip delay	0.270 - 0.390	0.1807 (distributed) 1.8570 (centralized)	sec sec
Jitter		0.1920 (distributed) 1.3621 (centralized)	sec sec
Data loss		0.28% (distributed) 0.97% (centralized)	- -

Table 4.17: Comparison of satellite communication system simulation to estimated and measured Iridium values.

"Iridium Drawl" experienced by Iridium users.[78]

If the routing architecture is switched to centralized, with all else equal, then the simulation produces an average RTD of 1.8570 seconds, a jitter of 1.3621 seconds and a loss of 0.97 percent.

According to McMahon and Rathburn,[79] voice latency over Iridium was measured to be 270-390 milliseconds, which is about twice the average round-trip delay estimated by the simulation under the first set of conditions and considerably less than that estimated under the second set. This result is very interesting because the Iridium routing architecture used regional networks for call service, which behaves as a combination of distributed and centralized routing. The actual Iridium routing architecture is distributed over many regional gateways, which act as centralized routers within the region. Thus, the results would match our expectations for the quality of service metrics.

There are a number of assumptions and differences, which could explain any further discrepancies between the QoS metrics and the measured Iridium values. The simulation only looks at a single point in time so the "jitter" isn't due to fluctuations over time, but rather the standard deviation of all of the round-trip delays calculated throughout the system. The round-trip delay is a function of the probability of packet error (corruption of data), the number of dropped packets due to congestion and the propagation delay. The signal-to-noise ratio calculations assume no signal regeneration, so the data corruption estimates tend toward a worst-case bound. Circuit-switched systems can incur data corruption and propagation delay, but not congestion-dropped packets. However, since the

number of subscribers (and thus simultaneous users) are significantly below the saturation point of the system, congestion-dropped packets should not be a contributing factor. In real systems with ISL technology, "jitter" and dropped call effects can occur due to inaccuracy in the ISL-switching hardware and early generation operational protocols. Further quality-of-service problems can be incurred by non-optimal satellite-ground visibility and low-diversity constellation design issues.

Additionally, the effects of hand-off and tracking are not accounted for and there are generic (non-Iridium specific) estimates of the switching delays and other processing times. The Iridium routing and ARQ protocols are proprietary to the company, so it is unlikely that the simulated (standard) protocols completely match.

All in all, it is remarkable that the simulation-generated system outputs, QoS measurements and projected services match up so well with the real system. The simulation results are consistent enough to provide sufficient fidelity to demonstrate the Strategic Evolution of Systems methodology.

Chapter 5

Deployable Communication Networks for Manned Lunar Exploration

The following case study applies the Strategic Evolution of Systems framework to a low-fidelity example of deploying an augmenting series of communication networks designed to support manned lunar exploration. The study is intended to illustrate how to qualitatively define the energy for a physical non-market traded system that is **not** defined in terms of monetary units. The study also addresses the situation in which there is no option to maintain the current system; here, the exploration landing sites are assumed to be visited only once, requiring the deployment of new assets during each mission stage. The case study fidelity may be refined in order to provide decision-maker support for identifying appropriate deployment strategies for future lunar missions.

NASA's Vision for Space Exploration[80] calls for manned exploration of the Moon by 2020 in preparation for manned Mars missions. The 2005 NASA development and exploration roadmap, shown in Figure 5.1, highlights the necessary technological development periods to enable these lunar and Mars exploration missions. According to the Vision for Space Exploration (VSE) and the roadmap, crewed vehicles are developed while robotic orbital and surface missions to the Moon are underway. Initial manned lunar exploration sortie missions, operationally similar to the Apollo J-type missions, include multiple excursions and Extravehicular Activity (EVA)'s in the area of the lander. Over multiple missions, permanently occupied outposts reminiscent of Antarctic research stations are developed.

The development and deployment of the communications infrastructure is missing from the roadmap

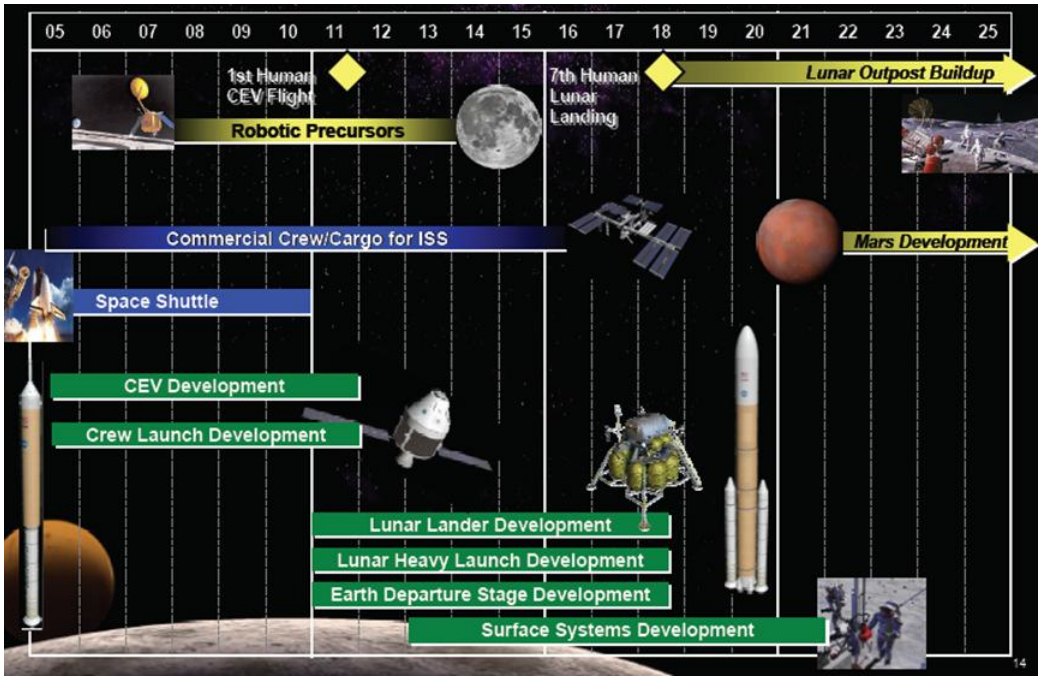


Figure 5.1: NASA development and exploration roadmap as of 2005. Picture from ESAS Final Report.[11]

depicted in Figure 5.1. Proposed communications architectures are presented in the Exploration Systems Architecture Study (ESAS)[11] and Concept Exploration and Refinement (CER) studies.[81] Carr[82] presents network quality of service analysis of several methods for deploying distributed ground-based relays for planetary surface explorations, including a basic Shannon measure of entropy metric used to measure directed divergence.

Any exploration architecture will fall somewhere between a purely ground-based and purely space-based architecture. The final architecture will be a hybrid of ground-based assets and space-based assets. This chapter will illustrate the application of the Strategic Evolution of Systems framework to address the following question at a high level: Given a minimalist initial combination of ground and space-based assets, what kind of communication deployment strategy has the strategic advantage?

5.1 Perspective

One of the purposes of the perspective portion of the framework is to ensure that the perspective of the client system is identified, acknowledged and appreciated immediately. A significant portion

of the information contained in this section is derived from and based on the Extensible Planetary Surface Mobility System study.[12]

5.1.1 Identify Objectives, Requirements and Constraints

This section describes the objectives, requirements and constraints for the lunar exploration communication infrastructure. The objectives and constraints flow from the desires of the stakeholders in the supported exploration system.

Objectives

According to the Extensible Planetary Surface Mobility Systems study, "the primary beneficiaries of Moon and Mars surface exploration are scientists (acquire data), explorers (access new locations), operators (gain experience) and the American public (enjoy 'armchair exploration' and sharpen interest in science and engineering)."

The stakeholder's desire access to valuable sites of interest and the acquisition of samples and information collected at those sites.

The objectives relevant to the communications infrastructure is the timely transmission of images, video and other data from locations and sites of interest back to Earth. Thus, the communication architecture is a value-delivery infrastructure. The exploration system delivers the value (information to be transmitted), while the communication system is a support infrastructure that enables this delivery (the physical system that transmits the information).

The objectives of the communication system include:

- Maximizing the percent access time enabled by the system. In other words, the availability of complete connections between Earth and the base and mobile assets.
- Maximizing coverage of the exploration area.
- Minimizing cost by minimizing launch mass.

In order to meet the communication systems objectives, it is necessary to identify the set of potential landing sites and the impact of visiting them from a communications standpoint. It is also important to understand the operational mechanisms — the Design Reference Missions — that will enable the collection of data to be transmitted.

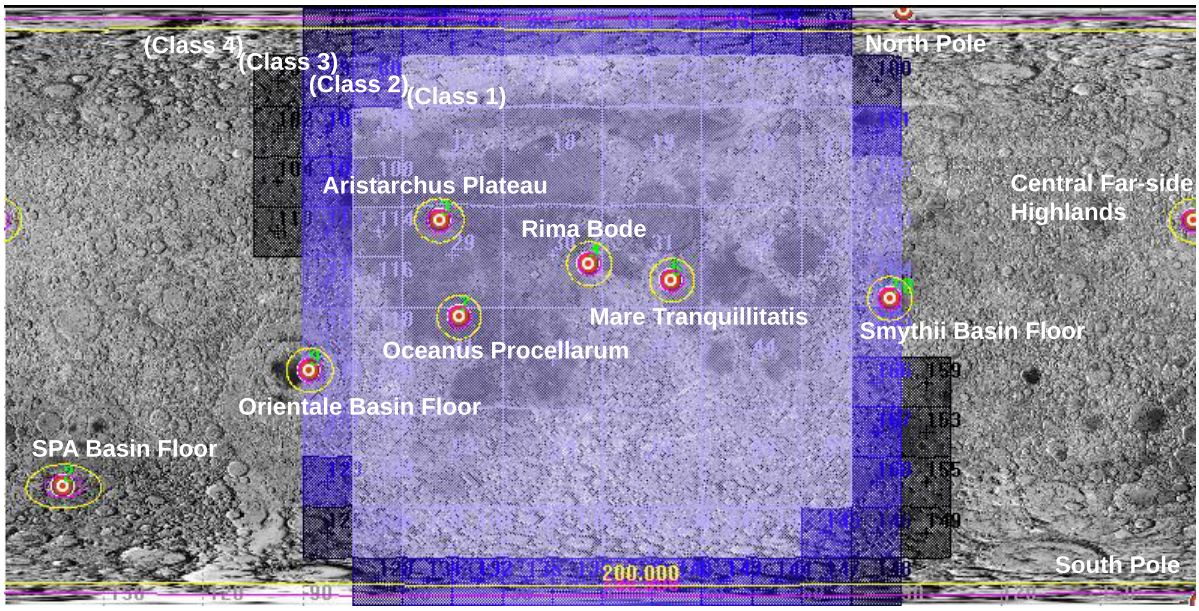


Figure 5.2: Map of lunar surface overlayed with Lunar Mission class regions designated by color (the lighter the overlay color, the greater the access) and identified near the top left of the figure. The landing sites are plotted and identified. Figure adapted from the Extensible Surface Mobility Systems study final report.[12]

Landing Sites: There are 10 potential landing locations (see Figure 5.2).[11] The first five sites are listed in Table 5.1 in the likely order of their exploration.

Given the potential landing sites, it is possible to find a basic measure of the direct communication coverage available between a communication asset at a given landing site and at least one Deep Space Network station. This impact is defined according to the landing site's Mission Class:[12]

1. *Continuous Direct Earth Coverage:* 97 percent or better coverage over one year.
2. *Cyclic Direct Earth Coverage, High Duty:* Repeating, non-continuous access with greater than or equal to 50 percent duty over one year.
3. *Cyclic Direct Earth Coverage, Low Duty:* Repeating, non-continuous access with less than 50 percent duty over one year.
4. *No Direct Earth Coverage:* No access over one year.

The duty cycles were estimated using Satellite Tool Kit (STK) access models assuming elevation angles of greater than 10 degrees, terrain grazing angles of greater than 5 degrees, and a lunar elevation angle of greater than 5 degrees. Figure 5.2 depicts the coverage map for the lunar surface. The lighter the color, the higher the duty cycle.

Site	Coordinates	Mission Class	Description
Mare Tranquillitatis (north of Arago)	8° N, 21° E	1	Likely first mission. Near Apollo 11 landing site. Smooth maria terrain. Easy to access.
Oceanus Procellarum (inside Flamsteed P)	3° S, 43° W	1	Likely second mission. Mare terrain site.
Smythii basin floor (near Peek)	2.5° N, 86.5° E	2	Likely third mission Mare terrain. Sometimes out of Earth view, so requires relay for continuous communications.
Central Far-side highlands (near Dante)	26° N, 178° E	4	Likely fourth mission. Terrain unclear. Would require relay satellites.
South pole (rim of Shackleton)	89.9° S, 180° W	4	Likely fifth mission. Terrain unclear.
SPA basin floor (near Bose)	54° S, 162° W	4	Would require relay satellites. Terrain unclear.
Aristarchus Plateau (north of Cobra Head)	26° N, 49° W	1	Terrain unclear.
Rima Bode (near vent)	13° N, 3.9° W	1	Terrain unclear.
North pole (rim of Peary B)	89.5° N, 91° E	4	Terrain unclear.
Orientale basin floor (near Kopff)	19° S, 88° W	2	Highland/mare terrain. Sometimes out of Earth view, so requires relay for continuous communications.

Table 5.1: Possible Lunar landing sites. Table based on information contained in the ESAS Final Report.[11]

Design Reference Missions: The Design Reference Missions (Design Reference Missions (DRM)) are the operational mechanisms that will enable the collection of material and information and were designed in the Extensible Planetary Surface Mobility Systems study "to cover the broad range of operational tasks that astronauts would perform on the Earth, Moon and Mars." For surface stays up to one week, astronauts live in the lunar lander and perform sortie missions. For longer term stays up to 180 days, the astronauts will live in a habitation module as part of a long-term outpost.

The DRMs include short distance excursions (DRM-1), long-distance excursions (DRM-2), re-supply logistics (DRM-3) and infrastructure operations (DRM-4). This case study will focus on DRM-1 operations. DRM-2 operations involve driving the camper about 100 kilometers (km) out from the main base and then setting up a temporary camp and executing DRM-1 excursions in the immediate area. The effect of DRM-2 operations on the results of this case study would be very interesting and should be considered for future study. Re-supply and infrastructure operations occur near to the base (maximum 2 kilometers), and need not be considered here.

The immediate vicinity around the lander or outpost is assumed to be explored using the DRM-1 short distance excursions. The operational format of DRM-1 missions is similar to the Apollo 15-17 Lunar Roving Vehicle (LRV) excursions. The astronauts on DRM-1 excursions are assumed to wear space suits while traveling on unpressurized vehicles.

Short distance excursion tasks include deployment of science instruments, collecting samples, con-

ducting detailed surveys, and photo and video documentation of the experience and site of interest.

The duration of a DRM-1 excursion is 8 hours, based on current space suit life-support capacity. The range is assumed to be 20 kilometers. The increase in range over the 11 kilometers traveled by the Apollo 15-17 astronauts away from the Lunar Excursion Module (LEM)[83] is an assumption based on the longer mission duration, the increase in crew and the likely increase in exploration area.

Requirements

This section discusses the high-level requirements for the communications infrastructure to support the planetary surface mobility system.[12] The requirements are separated into "hard" requirements, or those requirements that **must** happen, and "soft" requirements, or those requirements that **should** happen.

- **"Hard" Requirements:**
  - *Must transport information collected by the mobile assets to Earth at some point:* The communication infrastructure must ensure retrieval of the scientific data, a primary value-generator of the exploration system.
  - *Must provide continuous reliable communications between the base and mobile assets regardless of the line-of-sight:* This requirement is intended as a safety measure since there is no guarantee of communications between the mobile assets and the operational support based on Earth. Bases established on the near side of the Moon should always have direct line-of-sight with the Earth, but local terrain variations may prevent the mobile asset from the same. Exploration on the far side of the Moon will require some level of space asset deployment since neither the base nor the mobile assets will ever have direct line-of-sight with the Earth. It is expected that some astronauts will remain at the base during short-distance excursions and could provide basic support during contingency operations.
- **"Soft" Requirements:**
  - *Should provide continuous transmission of information between the mobile assets and Earth:* "Real-time" voice and telemetry communications and transmission of scientific

data is highly desirable, but is often impractical without significant infrastructure deployment to overcome orbital and terrain obstructions.

- Should support the command and control, telemetry and data collection of all of the mission elements for the duration of the mission lifetime.
- Should be flexible and evolvable to meet the demands of uncertain mission objectives, growth and complexity.

Implications for Asset Deployment: The high-level communication requirements have several implications for the deployment of ground- and space-based assets. It is always assumed that there are base and mobile elements on the ground.

There are two independent conditions that **do not require ground-based or space-based assets**. These independent situations are:

- The landing site and exploration area fall within Lunar Mission Class 1. This situation mandates that either (1) the base and mobile assets maintain line-of-sight contact or (2) both the base and mobile assets maintain direct communications with Earth but not necessarily with each other.
- The landing site and exploration area fall within Lunar Mission Classes 2 or 3 with "soft" communication requirements. This situation mandates that the base and mobile assets maintain line-of-sight contact.

The situations requiring neither ground-based nor space-based assets assume that the terrain conditions and elevation angles to the Earth are sufficient to enable line-of-sight access throughout the exploration area. Terrain diffraction can be a serious issue since the received signal strength is halved when the top of the terrain is at least as high as the line-of-sight path.

The use of a ground-network enables communications redundancy, ground-based navigational capabilities, and an increase in scientific payload and sample capacity on the mobile assets by shifting the majority of the communications equipment to the base. The subsequent increase in the gain of the communications equipment at the base would enable higher-throughput communications between the base and Earth. Thus, it makes sense to always deploy a ground-based network even if one is not strictly required.

There are three independent conditions that **require a ground-based network** (see Figure 5.3a) but do not require space-based assets:

- The landing site and exploration area fall within Lunar Mission Class 1 with no line-of-sight or direct communications constraints.
- The landing site and exploration area fall within Lunar Mission Classes 2 or 3 with "soft" communication requirements but no line-of-sight constraint.
- The landing site establishes a near-side communications base and enables border far-side missions.

Finally, there are two independent conditions that **require both a ground-based network and space-based assets** (see Figure 5.3b):

- The landing site and exploration area fall within Lunar Mission Class 4.
- The landing site and exploration area fall within Lunar Mission Classes 1, 2 or 3 with "hard" communication requirements.

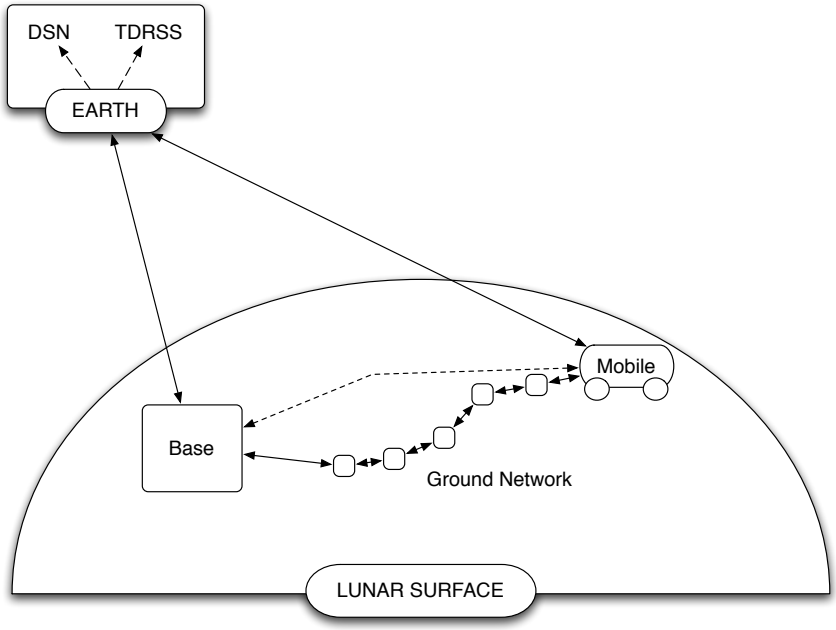
Constraints

There are two obvious physical constraints: the deployable ground-based assets must fit within the storage capacity of the transportation units (space and ground) and must not exceed mass budget limitations.

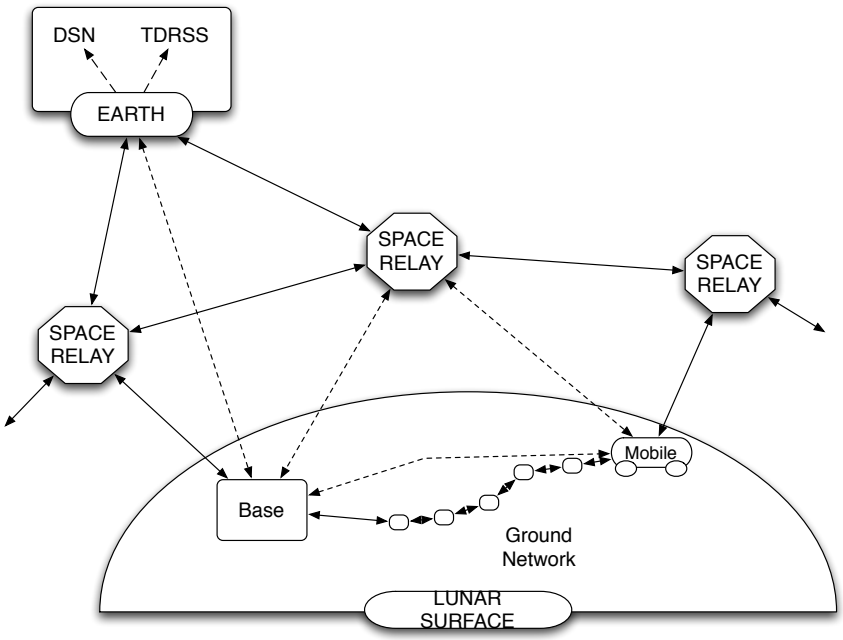
If there is no ground-based Local Access Network (LAN), then there must either be a direct line-of-sight connection between the mobile asset and the base, or 100 percent connectivity along the mobile-to-satellite-to-base link, or 100 percent connectivity between the mobile asset and Earth and Earth to base. This constraint is required to ensure the 100 percent availability between mobile and base requirement.

5.1.2 Identify Decisions, Bounds and Logical Constraints

This section describes the relevant decisions, the bounds on those decisions and the logical constraints between them.



(a) Architecture with Ground Relays



(b) Architecture with Ground and Space Relays

Figure 5.3: Architectural possibilities for (a) a ground-based architecture with ground relays and (b) a full architecture with ground and space relays. Earth assets include the NASA Deep Space Network and the Tracking and Data Relay Satellite System. Figure adapted from illustrations in Extensible Planetary Surface Mobility Systems study.[12]

Morphological Matrix of Lunar Communication Deployment Decisions								
Category	**Variable**	**Decision**	**Units**	**Alt A**	**Alt B**	**Alt C**	**Alt D**	**Alt E**
Space	AddC	Add Lunar Satellite Plane	[-]	Yes	No			
Space	CIncl	Orbital Inclination	[deg]	27	50	56.2	76	86
Space	CEcc	Orbital Eccentricity	[-]	0	0.6			
Space	CAlt	Altitude at Semi-major Axis	[kilometer (km)]	2,737	6,541			
Space	CRaan	Right Ascension of Ascending Node	[deg]	0	90	180	270	
Space	CPer	Periapsis	[deg]	0	90			
Ground	AddLAN	Add Excursion LAN Network	[-]	Yes	No			
Ground	LANType	Type of LAN Product Platform	[-]	VSS	MSS	LSS		
Strategy	Deploy	Relay Deployment Strategy	[-]	Straightline	Adaptive			

Table 5.2: Morphological matrix of decisions for the staged lunar deployment illustration.

Logical Constraints for Lunar Communication Deployment Decisions		
Name	**Scope**	**Equation**
Plane Polar	AddC, CIncl, CEcc, CAlt, CRaan and CPer	$(addC == Yes)$ && $(CIncl == 56.2deg)$ && $(CEcc == 0.6)$ && $(CAlt == 6,541km)$ && $(CRaan == 0deg)$ && $(CPer == 90deg)$
Plane 27	AddC, CIncl, CEcc, CAlt, CRaan, CPer	$(addC == Yes)$ && $(CIncl \neq 27deg)$ && $(CEcc == 0)$ && $(CAlt == 2,737km)$ && $(CRaan == 0deg)$ && $(CPer == 0deg)$
Plane 50	AddC, CIncl, CEcc, CAlt, CRaan, CPer	$(addC == Yes)$ && $(CIncl \neq 50deg)$ && $(CEcc == 0)$ && $(CAlt == 2,737km)$ && $(CRaan == 90deg)$ && $(CPer == 0deg)$
Plane 76	AddC, CIncl, CEcc, CAlt, CRaan, CPer	$(addC == Yes)$ && $(CIncl \neq 76deg)$ && $(CEcc == 0)$ && $(CAlt == 2,737km)$ && $(CRaan == 180deg)$ && $(CPer == 0deg)$
Plane 86	AddC, CIncl, CEcc, CAlt, CRaan, CPer	$(addC == Yes)$ && $(CIncl \neq 86deg)$ && $(CEcc == 0)$ && $(CAlt == 2,737km)$ && $(CRaan == 270deg)$ && $(CPer == 0deg)$

Table 5.3: Logical constraints for the staged lunar deployment illustration, defining the available satellite constellations.

The high-level decisions are shown in Table 5.2. The decisions are broken down into whether they apply to the space-based assets (the satellite constellations) or the ground-based LAN networks. If it is decided that a satellite plane will be augmented to the network, then the characteristics of that constellation must be specified using the orbital inclination, the orbital eccentricity, the altitude at the semi-major axis, the Right Ascension of Ascending Node (RAAN) and the periapsis. If a LAN is added, then the type of relay must be chosen as well as the strategy used to deploy them. For simplicity, the relays are assumed to be carried and deployed from the mobile asset.

The logical constraints shown in Table 5.3 define the satellite constellations. The South Pole orbit is recommended in *A new Paradigm for Lunar orbits*.[84] The other inclinations are known frozen orbits for the moon. These orbits are rare due to the highly uneven gravity field.

The LAN relays are assumed to come in three product platform variations: VSS (very small station disposable relay), MSS (medium size station infrastructure-building relay) and LSS (large size station established infrastructure relay).

The deployment strategies are straightline drop and adaptive deployments using the mobile asset to carry and distribute the relays. The straightline drop strategy is the simplest to implement and model. The mobile asset is assumed to travel in a straight line across the terrain. When the mobile asset loses its connectivity to any relay, it drops another relay immediately behind it. The adaptive strategy is similar. However, when the mobile asset senses a connection loss, a new relay is placed at the highest elevation point within a pre-determined radius.

There are a number of assumptions concerning the design and sizing of the mobile asset, base, relay, and Earth-based and satellite assets.

Based on the Extensible Planetary Surface Mobility Systems study, the mobile asset is an Unpressurized Vehicle (UPV) sporting a 0.5-meter diameter antenna installed at a height of 2-meter above the lunar surface. The CER study sizes this antenna such that the transmit power is 30 Watts (W).

The base antenna is assumed to be placed on top of the lunar lander at a height of 9-meters. The 4-meter diameter, Ka-band, 200-Watt (W) antenna concept is based on the moon-base antenna design proposed in the CER study.

The VSS are assumed to be little balls with a diameter of approximately 0.1-meter (the antenna height is thus about 0.1-meters), while the MSS and LSS are assumed to be 1-meter and 2-meters in height, respectively.

Based on the CER study, the DSN stations on Earth are assumed to be the 2-kW Ka-band, 34-meter antennas subnet. Thus, these DSN antennas act as the legacy system. The other well-established option is the Tracking and Data Relay Satellite System (TDRSS) satellite system.

In future study, the power budgets of the assets and the satellites should be sized to meet the 1 Megabits per second data rate requirement determined by the CER study.

5.1.3 Generate Relevant Architecture Instantiation Networks

Next, the transition and evolution rules for the system must be determined in order to generate the relevant architecture instantiation networks.

Transition Rules

For simplicty, only one transition rule will be considered. The landing sites are assumed to be visited once; the deployment of communication assets act as augmentations to the existing network

rather than transitions in the sense discussed up until this point in the thesis.

The augmentation transition rule thus specifies that each satellite plane can only appear once along an option path; a similar rule exists for the LAN networks.

Evolution Rules

The evolution rules are designed to simplify the option space and to ensure that the communication requirements are met.

It is assumed that there are 10 deployment stages, one for each landing site. Each stage represents a new exploration mission at a new site.

The first five stages are assumed to follow the order of the first five landing sites listed in Table 5.1. Thus, the first stage is assumed to be a mission to Mare Tranquillitatis, the second stage to Oceanus Procellarum and so on. The landing sites for the remaining five stages are assumed to be unknown but bounded to the five unvisited sites of interest. Since these sites can be visited in any combination, this assumption results in 120 scenarios.

According to the system requirements, there must always be a path through the communication network connecting the Earth-based assets to all surface assets. Any option path that violates this rule is eliminated from consideration.

5.1.4 Estimating the Transition Costs

Since the deployment of lunar communication networks is an augmenting system, the transition costs are simply the cost of deploying the additional assets at each stage. Please refer to the resource estimation section and Table 5.6 for the results of these cost estimations.

5.1.5 Estimating the Architecture Desirability

In this case study, the costs and benefits of the systems, and therefore their desirability, are a bit different to calculate since the benefits cannot be described in terms of monetary units. Thus, this example tends more toward cost-effectiveness for the energy calculations than toward benefit-cost.

Costs are measured in terms of the resources expended to achieve these benefits. Benefits are measured in terms of the value delivery enabled by the communication system.

In this example, the costs and benefits are not monetary in nature, thus the meaning of E = 1 must be redefined. For monetary systems, E = 1 represents the pay-off point of the system, when the benefits accrued are equal to the resources expended. Here, E = 1 will be defined as the point at which the staged deployment architecture is as desirable as the benchmark of launching a full lunar constellation with a pre-deployed ground network. Thus, a sub-unity energy means the staged deployment architecture is preferable to the comparison architecture.

The energy of the network deployed upfront can be written as follows, where α is the conversion constant:

$$E_B = \alpha \frac{C_B}{B_B} = 1 \tag{5.1}$$

$$\alpha = \frac{B_B}{C_B} \tag{5.2}$$

All of the energies calculated for the comparison strategy must be adjusted by α.

Resource Estimation

To first order, the resources expended to achieve a certain level of benefit can be estimated as the launch mass of those resources, as this is one of the biggest cost drivers in space exploration systems. The two types of assets under consideration are the in-orbit satellite constellations and the surface relays.

Mass estimation for the relays is complicated by the fact that it is dependent on the type of relay (VSS, MSS, LSS) and the number of relays required.

Modeling the Launch Mass of the Satellites: This section briefly describes the estimation of the satellite launch mass used in this low-fidelity case study.

The comparison full-global constellation is based on the constellation designs proposed by Ely and Lieb.[85] According to Ely and Lieb, the six spacecraft in their design would weigh 1,075 Kilograms each. Thus, the launch mass of the satellites is a combined 6,450 Kilograms.

Let it be assumed that there are six 12-meter antennas per longitude. There are 10 landing sites, so it shall be assumed there need to be 10 times six 12-meter antennas on the ground. If each of

these antennas weighs 100 Kilograms, this adds another 6,000 Kilograms. The upfront resource expenditure for this architecture is thus estimated to be 12,450 Kilograms.

For the augmenting satellite planes, assume that their mass is 900 Kilograms. They would be similar to the spacecraft proposed by Ely and Lieb but the lower altitude would mean smaller spacecraft. Future work should use the link budgets to estimate the antenna size and power, and thus mass. Since each constellation is assumed to have five satellites, this gives a resource penalty of 4,500 Kilograms per plane.

Modeling the Launch Mass of the LANs: This section briefly describes the estimation of the LAN launch masses used in this low-fidelity case study.

For the deployable ground relays, an estimate of the launch mass per relay as well as the number of relays must be made. To estimate the number of required relays per strategy, it is necessary to get a sense of how many relays are required to guarantee connectivity for different types of terrain. This can be done by estimating the distance traveled before another relay must be placed for different types of terrain.

This case study focuses on the DRM-1 missions only. The specifics of the LAN deployment for these missions are strongly influenced by the terrain. In the Extensible Planetary Surface Mobility Systems study, a monte carlo analysis of the effect of terrain uncertainty on the number of relays required to guarantee connectivity between the base and mobile asset was performed. The results of the analysis are summarized below and are used to estimate the launch mass resource penalty.

Four types of terrain — smooth mare, hummocky upland, rough mare and rough upland terrain — were modeled using a power spectral density method.[12] Regions of representative terrain were generated for each terrain type. The vehicle is simulated deploying a communications relay and moving east in a straight line along the representative terrain. The vehicle checks for line-of-sight connectivity with the deployed relay every 2-meters; when the connectivity is lost, a new relay is deployed according to the specified deployment strategy.

An example output of this process is demonstrated in Figure 5.4 and Figure 5.5. In this case, the vehicle traverses Hummocky Upland terrain deploying 1-meter MSS relay antennas using the straight-line deployment strategy. Here, a total of 12 relays were required to guarantee connectivity over a distance of 270-meters between the first and last relays. The average distance per relay is 22.5-meters, implying the need for 44 relays to cover a 1 kilometer traverse.

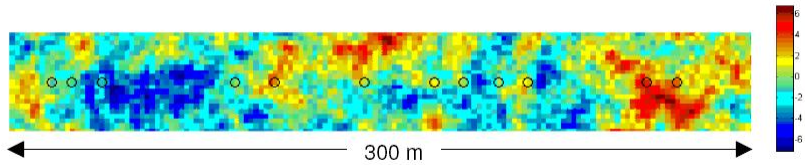


Figure 5.4: Example placement of relays for straightline deployment over Hummock Upland terrain with 1-meter MSS relay antennas. The colour scale indicates the elevation from nominal (+6-meters to -6-meters). This figure was taken from the Extensible Surface Mobility Systems study final report.[12]

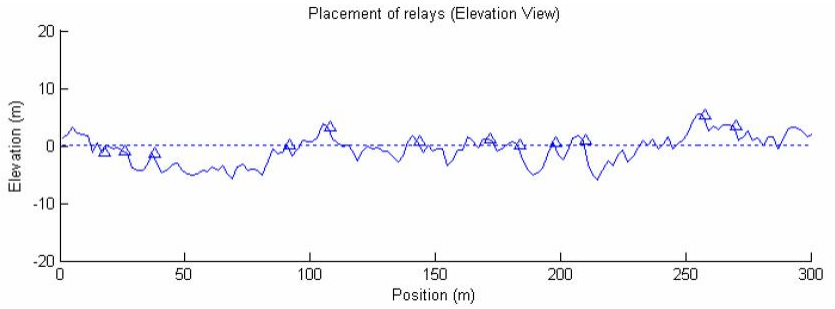


Figure 5.5: Corresponding two-dimensional elevation map for the straightline deployment relay placement. This figure was taken from the Extensible Surface Mobility Systems study final report.[12]

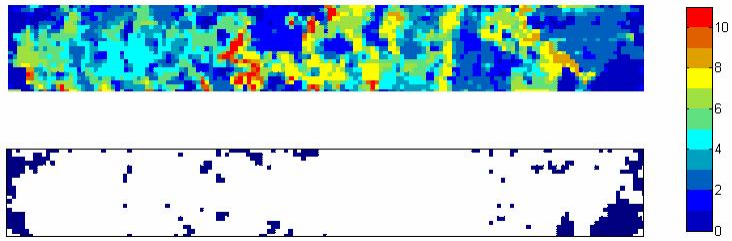


Figure 5.6: Corresponding connectivity maps for the straightline deployment relay placement. This figure was taken from the Extensible Surface Mobility Systems study final report.[12]

Figure 5.6 shows the connectivity maps for the aforementioned example traverse. The top image depicts the number of relays visible from each point in the 300-meter by 40-meter region, providing a measure of the overlap and robustness of the network as seen from the lunar surface. The bottom image illustrates the connectivity/no connectivity areas. The white areas have connectivity, the blue areas do not. The relay network has fairly good coverage of the lunar surface in the area, providing opportunities for "in situ" monitoring systems. Vehicle and astronaut antennas would be even more visible to this network as their antenna height is significantly above the lunar surface.

The connectivity maps enable first-order estimations of "in situ" sensor coverage statistics for the exploration area (the ratio of visible area to total area), as well as the inherent redundancy of the architecture. The greater the diversity at the location of a relay, the more robust the system is to random failures. Extending this work should provide inclusion of this measurement of coverage and redundancy.

The difficulty with modeling the launch mass of the LANs is that it is dependent on where they will be deployed as well as the number of excursions. The Un-Pressurized Vehicles (UPVs) were designed assuming an operational range of 60 kilometers, though the Mission Utility Simulation Environment (MUSE) simulation environment in the Extensible Planetary Surface Mobility Systems study found that the UPV only traveled an average of 30 kilometers per excursion.[12] Furthermore, it was found that an average of 5.71 sites were visited per DRM-1 excursion. To first order, assume that the distance traveled per excursion is 30 kilometers, though if the drive back is constrained to parallel the original path out such that the vehicle maintains connectivity with the relay network, the distance covered by relays can be cut in half. With sufficient fidelity, the estimate of the number of relays per excursion can be found depending on the expected terrain type. Furthermore, assume that there is no overlap of the relay coverage for the set of excursions at each landing site. This assumption gives a worst case estimate for the required number of relays. The Extensibility study assumes five excursions per sortie mission (week-long stays) and 60 excursions for outpost missions (180 day stays). Only sortie missions will be considered.

There are several natural constraints that occur. The first is that the mass and volume of the relays for each excursion must fit within the payload capacity of the two UPVs that go out on the excursion (a total of 0.2 cubic meters and 60 Kilograms based on Extensibility study). The second is that the mass and volume of all of the relays used on the mission must fit within the payload of the lander. Of course, this can be relaxed by sending a separate module containing the relays. Either way, using launch mass as the resource metric means that the particular strategy does not matter for the purposes of this example.

The design of the relays are simplistic for illustration purposes. Table 5.4 summarizes the discussion that follows.

The VSS relays are assumed to be little balls with a diameter of 0.1-meters. Each VSS then takes up approximately 0.00052 cubic meters. It is assumed that the VSS relays weigh on the order of a cell phone (about 0.15 Kilograms). Thus, given the design of the UPV and its sample capacity, there

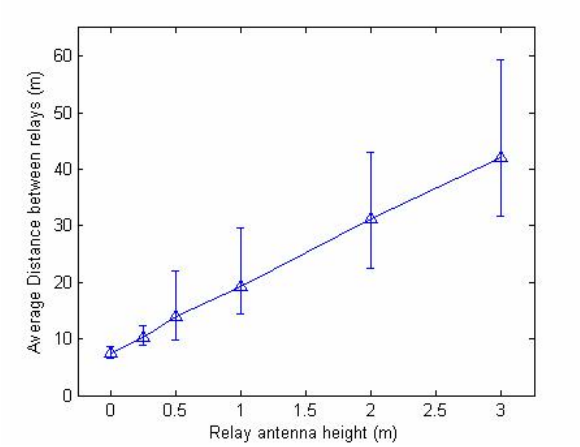


Figure 5.7: Range of relay numbers as a function of antenna height. This figure was taken from the Extensible Surface Mobility Systems study final report.[12]

Ground Relay Assumptions			
Relay	**Mass [Kilogram (kg)]**	**Volume [m^3]**	**Meters of Coverage per Relay**
VSS	0.15	0.00052	10
MSS	0.3	0.001	20
LSS	0.45	0.0015	30

Table 5.4: Ground Relay Assumptions for VSS, MSS, and LSS.

can be a maximum of 192 VSSs stored on each UPV per excursion (a total of 384 per excursion). For the time being, however, the constraint will not be enforced. Instead, it will be educational to see the impact of this communications architecture on the vehicle design.

It will be assumed that the MSS will have an antenna diameter of 0.1-meter but will extend to a height of 1-meter. Since there is no atmosphere on the moon, and therefore no wind, it may be possible to construct a very low-weight antenna station to achieve these dimensions. Thus, it will be assumed that the mass of the MSS is twice the weight of the VSS and 1.5 times again for the LSS. Similarly, let it be assumed that the volume is of the same scale (scaled based on number of relays).

Based on Figure 5.7, it can be assumed that for the VSS, it will take about 10 meters per relay for coverage, 20 meters per relay for the MSS, and 30 meters per relay for the LSS. This is a very rough estimate based on an extremely basic decision metric for deployment.

Given these assumptions, there is effectively no difference between VSS, MSS and LSS. For simplicity, consider only the VSS deployment. For five 15-kilometer excursions, it is expected that there will be about 7,500 VSS relays per mission. This results in a resource penalty of 1,125 Kilograms

per site with a LAN.

This resource penalty is a worst-case estimation. The relay mass is a function of the number of relays required to guarantee line-of-sight connectivity, which is determined to exist in MATLAB if the straight-line path between two points is free of obstacles. There is no connectivity if an obstacle touches or otherwise interrupts the line-of-sight path. In reality, diffraction enables signals to propagate over obstacles at the expense of reduced received energy levels. Obstacles below the line-of-sight path result in received energies at approximately the same strength as the transmitted energy. When the obstacle touches the line-of-sight path, the received signal strength is reduced by half. The power loss increases as the line-of-sight path is further obstucted by the obstacle.[86]

The range of the relays could be increased by taking advantage of diffraction. The greater the relay range, the fewer relays would be required to guarantee connectivity and the lower the resource penalty. Leveraging diffraction losses would require increased signal power from the relays to overcome the drop in received energy.

Modeling Value-Delivery

In this case study, the value delivery is assumed to be a function of the access time between Earth and the base enabled by each architecture option. For simplicity, it is assumed that the LAN connectivity is automatically 100 percent.

When a satellite constellation enables access between Earth (via DSN) and a landing site, it can be assumed that to first order, the requirement that the base and Earth be in contact has been satisfied. The degree of satisfaction of that requirement can be measured by the percent access over a year. This does not necessarily mean satisfaction of the hard requirement that the mobile and base be in continuous contact as there may not be sufficient line of sight between the mobile and any one of the DSN stations or any of the overhead satellites. Continuous communications between mobile and base requires that there be line of sight between the mobile and a relay asset as well as between a relay asset and the base (with contact between the two relay assets).

For the purposes of this case study, we shall assume that the percent access will also apply to the mobile-base element if there are no ground relays. This assumption will need to be relaxed in further studies.

The percent access over a sample time between the DSN stations and the potential landing sites via

Table of % Access Time Between DSN and Landing Sites Via Constellation Options									
Landing Site	**Site Direct**	**Plane 27**	**Plane 50**	**Plane 76**	**Plane 86**	**Plane Polar**	**"Full" Total**	**"Full" Plane A**	**"Full" Plane B**
Mare Tranquillitatis	1	0.9337	0.4963	0.6847	0.1348	0.5875	0.9971	0.9719	0.9971
Oceanus Procellarum	1	0.9352	0.4695	0.4430	0.3694	0.2819	0.9971	0.9309	0.9971
Smythii Basin Floor	0	0.8832	0.9344	0.0078	0.7587	0.7728	0.9971	0.9659	0.9967
Central Far-side Highlands	0	0.7763	0	0.8399	0.0479	0.3606	0.9845	0.9532	0.4940
South Pole	0	0	0.2154	0.9956	0.9958	0.9595	0.9971	0.9247	0.7570
SPA Basin Floor	0	0.0491	0.9949	0.9863	0.5629	0.8119	0.9912	0.7304	0.9783
Aristarchus Plateau	1	0.9927	0.9853	0.3270	0.4537	0.2437	0.9971	0.9588	0.9971
Rima Bode	1	0.9677	0.7214	0.7852	0.0013	0.4594	0.9971	0.9696	0.9971
North Pole	0	0	0.1909	0.9957	0.9958	0.0859	0.9971	0.7627	0.9318
Orientale Basin Floor	0	0.1231	0.7860	0.0912	0.8103	0.0570	0.9971	0.7148	0.9920

Table 5.5: Percent access time between the DSN stations and the landing sites via the constellation options. The comparison upfront deployment system ("full") is shown broken down by Plane A and Plane B and the combination of the two.

the various constellation plane options was found using STK. A table of these values is shown in Table 5.5. For comparison, the access times for the the upfront deployment constellation ("real") network are included.

As the system is augmented with multiple planes, the percent access numbers should increase. For simplicity, the case study will assume that the percent access value will become the maximum of the augmented values for any given site. Future studies can evaluate the actual impact of augmentation on the value-delivery metrics.

For the purposes of this case study, it shall be assumed that if a LAN is in place between the base and a mobile asset, then there is continuous communication between them, and the value of mobile-base communications is unity regardless of whether or not there is DSN-to-site access. This is not necessarily the case. The corresponding value of DSN-to-base and DSN-to-mobile connectivity will be the percent access value in the table, dependent on the situation.

The benefit metric is defined simplistically as the maximum network Earth-to-base access for the landing site visited in the current stage.

5.1.6 Options

The independent minimal states become: Plane Polar, Plane 27, Plane 50, Plane 76, Plane 86 and the VSS LAN. If all of the possible states are considered simultaneously, this would be 121 options per stage. The direct coverage option is always enabled.

Final Maximal Set of Options with Resource Penalty Estimate			
Option	**Description**	**Resource Penalty [kg]**	**Graph Color**
1	Add LAN at Location	1,125	Blue
2	Add Plane 27 + Add LAN at Location	5,625	Green
3	Add Plane 50 + Add LAN at Location	5,625	Red
4	Add Plane 76 + Add LAN at Location	5,625	Cyan
5	Add Plane 86 + Add LAN at Location	5,625	Magenta
	Benchmark Architecture	12,450	"origin"

Table 5.6: The final set of options for the lunar case study with an estimation of the resource penalty as a function of launch mass. The colors used to distinguish the options in the Strategic Advantage Maps are included.

Reducing the Option Space

To reduce the option space, assume that at most one plane and one LAN can be added in any given stage. This reduces the maximum number of options per stage to 12, but the number of resulting option paths are still quite large. There is no option to maintain the network "as is" since each site is visited only once. Furthermore, it is possible to eliminate the polar plane option as it is always dominated by another constellation regardless of the landing site. This reduces the number of options per stage to five. As the options are chosen, this number is reduced due to the evolution rule.

The final set of options in any given stage are shown in Table 5.6. The colors used to distinguish the options in the Strategic Advantage Maps are included.

5.2 Position

This section illustrates how the the system position is described and evaluated for the lunar network deployment example. The current energy of the system is found, the method used to estimate the terminal entropy values is discussed and the initial Strategic Advantage Maps are presented and explained.

5.2.1 Scenarios

The landing sites are assumed to be the sole source of uncertainty. Thus, the scenarios are combinations of the landing sites.

As mentioned during the discussion of the evolution rules, it is assumed that there are 10 deployment stages, one for each landing site. Each stage represents a new exploration mission at a new site. The first five stages are assumed to follow the order of the first five landing sites listed in Table 5.1. Thus, the first stage is assumed to be a mission to Mare Tranquillitatis, the second stage to Oceanus Procellarum and so on. The landing sites for the remaining five stages are assumed to be unknown but bounded to the five unvisited sites of interest. Since these sites can be visited in any combination, this assumption results in 120 scenarios. If every site other than Mare Tranquillitatis is unknown, the number of scenario combinations jumps to 2,880.

5.2.2 Initial Energy

This case study assumes that there is no infrastructure already in place on the lunar surface. At first glance, the initial energy appears to be undefined (0/0) since the Deep Space Network has no resource penalty because the assets are based on Earth and are therefore not launched. However, the legacy can be formulated as the Direct DSN access that is automatically available. It will be assumed that the cost of having this access is $0 according to the definition of the resource penalty.

The initial energy is therefore 0.

When the Mare Tranquillitatis site is visited in the first stage, the augmentation of the LAN increments the benefit by one and the resource penalty is incremented by the mass estimate of the launching and transporting the relays.

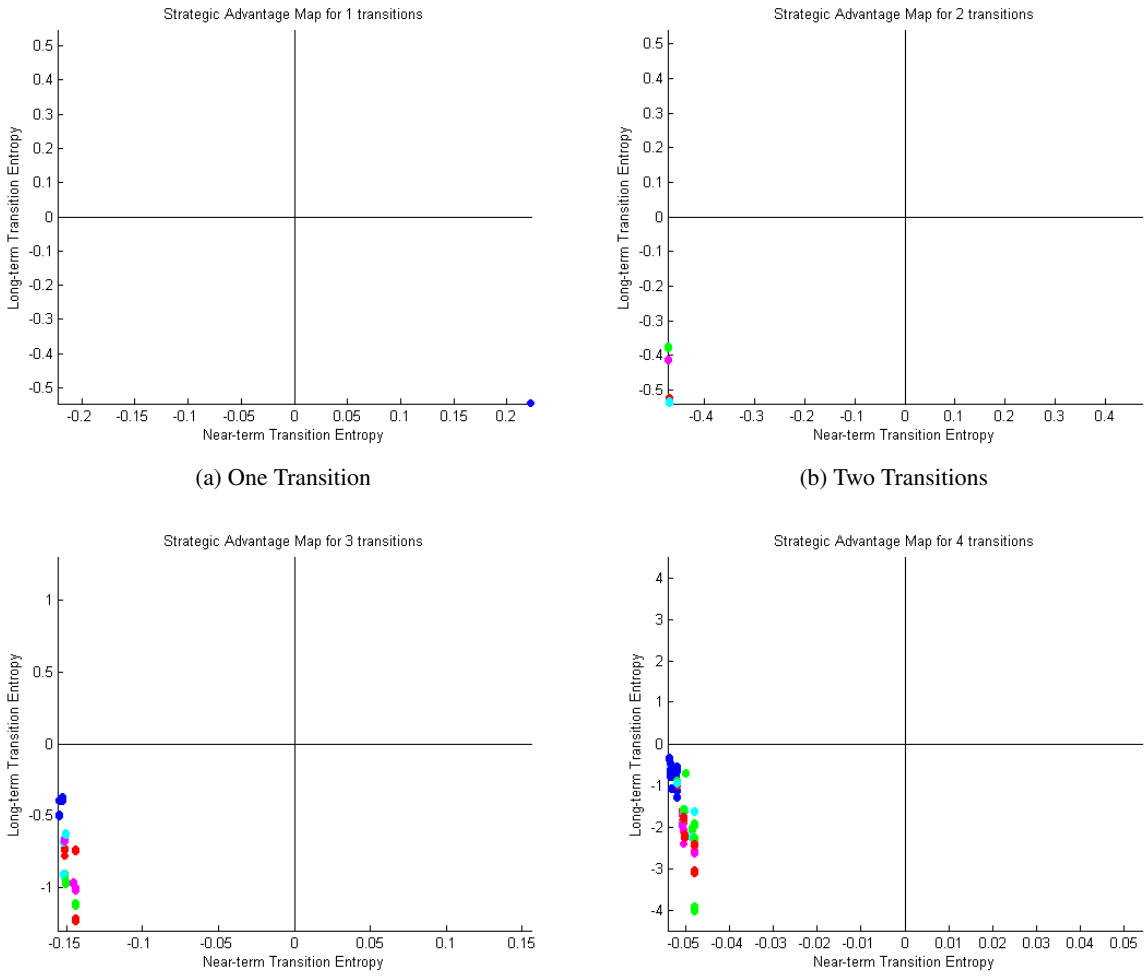
Thus, when Mare Tranquillitatis is visited, the energy becomes:

$$E = \alpha\left(\frac{1125}{2}\right) = 562.5\alpha$$

5.2.3 Estimation of Terminal Values

It is assumed that the benefits are not time-varying within an option selection and the resources are expended only once. Thus, the terminal entropy values can be estimated as the entropy (new entropy term) of the final-stage energy.

Further work can look at the impact of obsolescence in this system.



(a) One Transition

(b) Two Transitions

(c) Three Transitions

(d) Four Transitions

Figure 5.8: Strategic Advantage Map for Stage 1 looking forward four stages.

5.2.4 Initial Strategic Advantage Map

The legacy is assumed to be the Direct DSN access, and the first stage is assumed to land at Mare Tranquillitatus. The Initial Strategic Map in Figure 5.8 is thus looking out after adding the LAN deployment at Mare Tranquilliatus. The assumption is that since the second landing site is also in a Lunar Mission Class 1 region, only a LAN will be deployed and no satellite planes are required. Thus, the next move will automatically be to add a LAN to the Oceanus Procellarum site. The initial map is colored by the option transitioned to in the time period noted. To recap from Table 5.6, blue represents adding only LAN, green represents adding Plane 26, red adds Plane 50, cyan adds Plane 76 and magenta adds Plane 86. It is crucial to note that the origin is not a measure of

the legacy system since the option to maintain the system is not meaningful. This option fails the fundamental requirements of the system. Thus, the transitions are measured relative to zero entropy, and provide a measure of desirability relative to the comparison upfront launch of an entire global network architecture. Without information about the intermediary augmentations, it is difficult to say what the Transition 3 and the Transition 4 maps tell us, other than it looks unlikely that the staged options considered are preferable to upfront deployment of the entire system.

Transition 1 has a single point in Quadrant IV because it is assumed that the network is automatically augmented with another LAN at the second landing site. At this stage, the staged deployment architecture has the preferred near-term entropy but the long-term entropy clearly favors the upfront deployment comparison system. Transition 2 indicates that Add Plane 26 (green) option will be the preferable option for two stages out.

Decision for Stage 1: By the evolution rule, add a LAN to the Oceanus Procellarum site and prepare to deploy Plane 26.

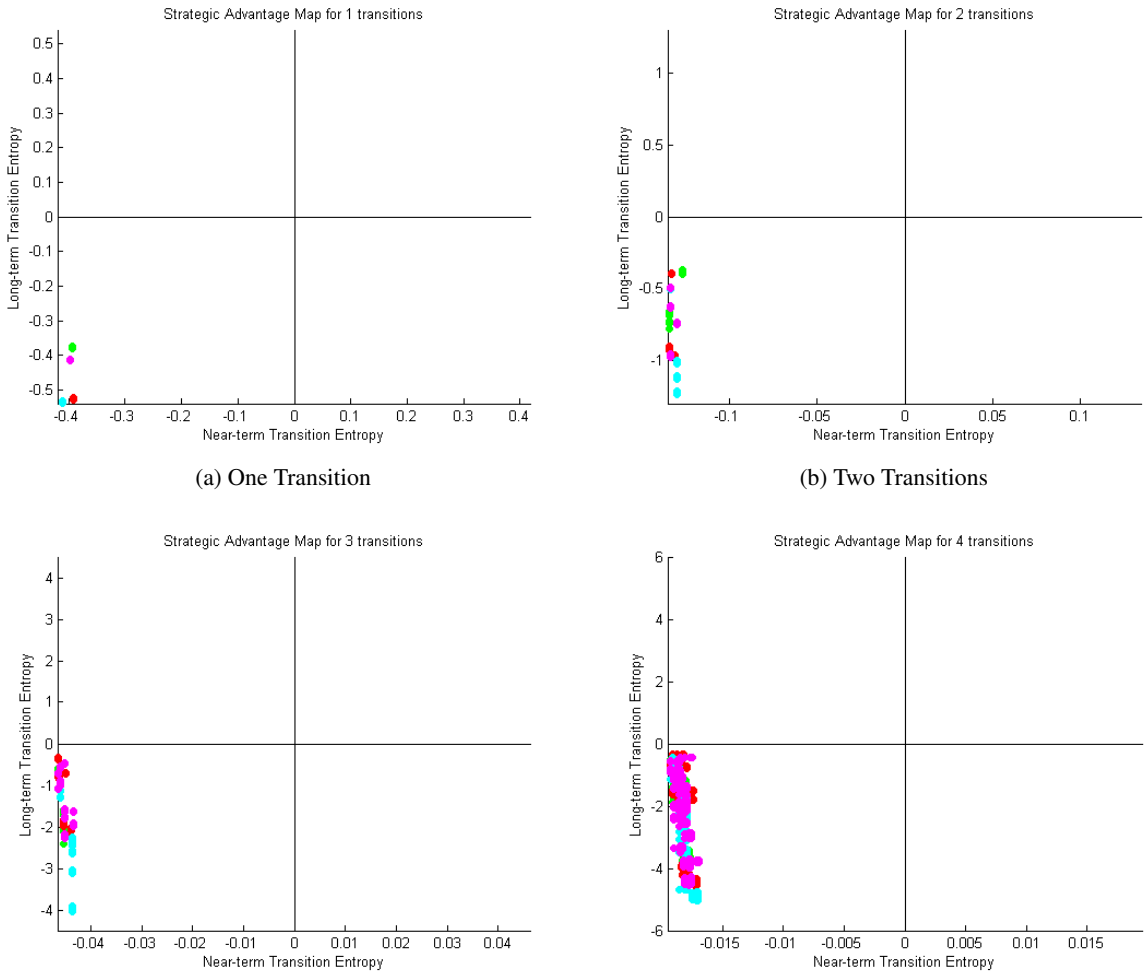
5.3 Results: Plan and Pattern

Plan and pattern require evaluating the position according to the perspective and using this information to decide on the best course of action as necessary while documenting the decision. For the purposes of this case study, plan and pattern will consist of simulating the series of decisions until the end of life by evaluating the current information, making a decision, documenting the reasoning and incrementing the time period.

5.3.1 Strategic Evolution

In this section, the scenario is propagated through time, with a "human in the loop" examining the output maps and making a decision as to the best course of action. This information is fed into the simulation and the next stage of the 'actual' landing site is propagated. The updated maps are the outputs and the next decision must be made. It is important to note that the horizon window fixed since there are only 10 landing sites to be visited in 10 stages.

As discussed in the previous section, the immediate transition should be to add a LAN to Oceanus Procellarum with preparations for deploying Plane 26. The outcomes of each time period are discussed below, with documentation on the reasoning for the next decision.



(a) One Transition

(b) Two Transitions

(c) Three Transitions

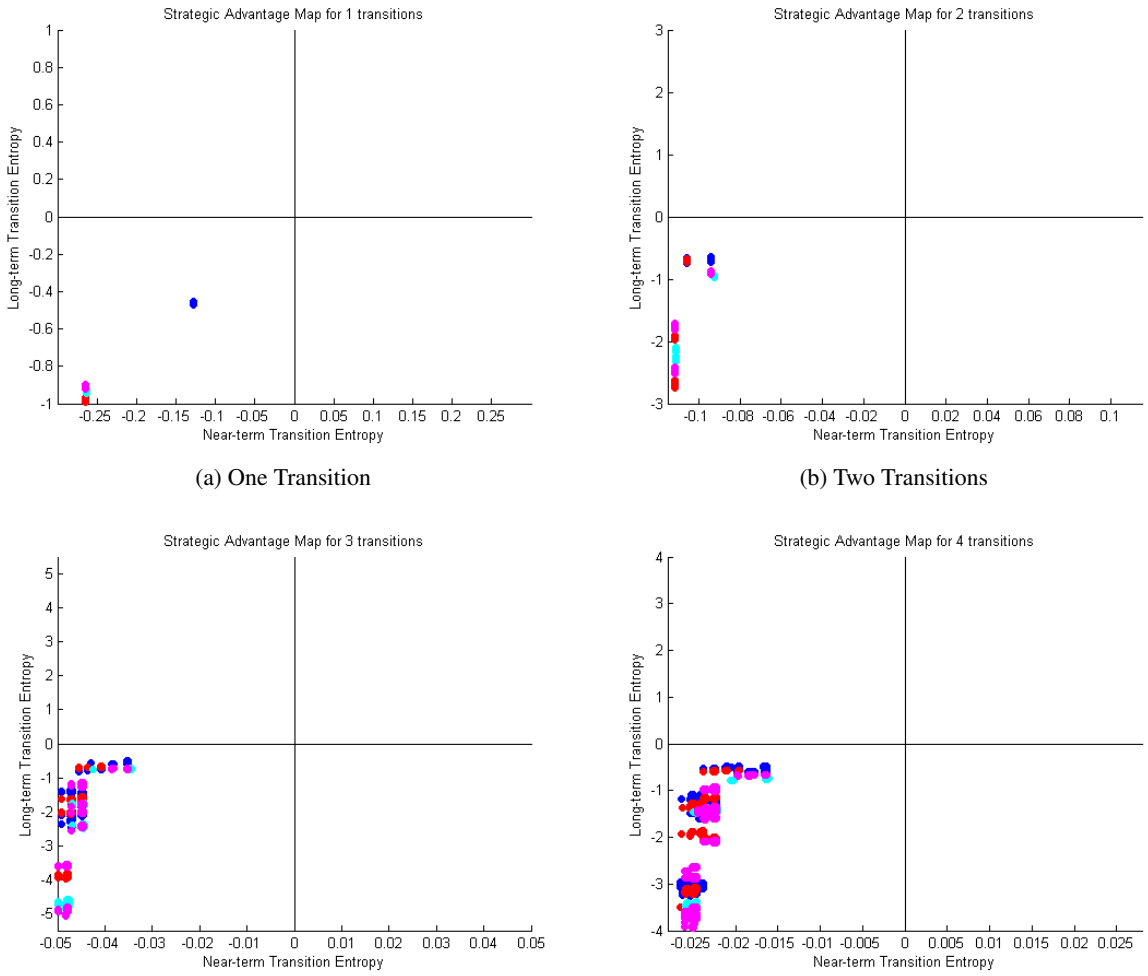
(d) Four Transitions

Figure 5.9: Strategic Advantage Map for Stage 2 looking forward four stages.

Stage 2

The results of Stage 2 are shown in Figure 5.9. This map and each of the following are colored by the immediate transition according to Table 5.6.

For the following two stages, the add Plane 26 option is preferable which implies that it is the best decision for these stages. However, the position of the add Plane 26 option drops in the third and fourth stage out, which reflects its drop in desirability due to the several sites it cannot access. Although it is difficult to see, there are green dots representing Plane 26 up near the optimal range for both stages. Thus, it would appear that adding Plane 26 is the best decision for the next two stages as well as providing good position for the later stages.



(a) One Transition

(b) Two Transitions

(c) Three Transitions

(d) Four Transitions

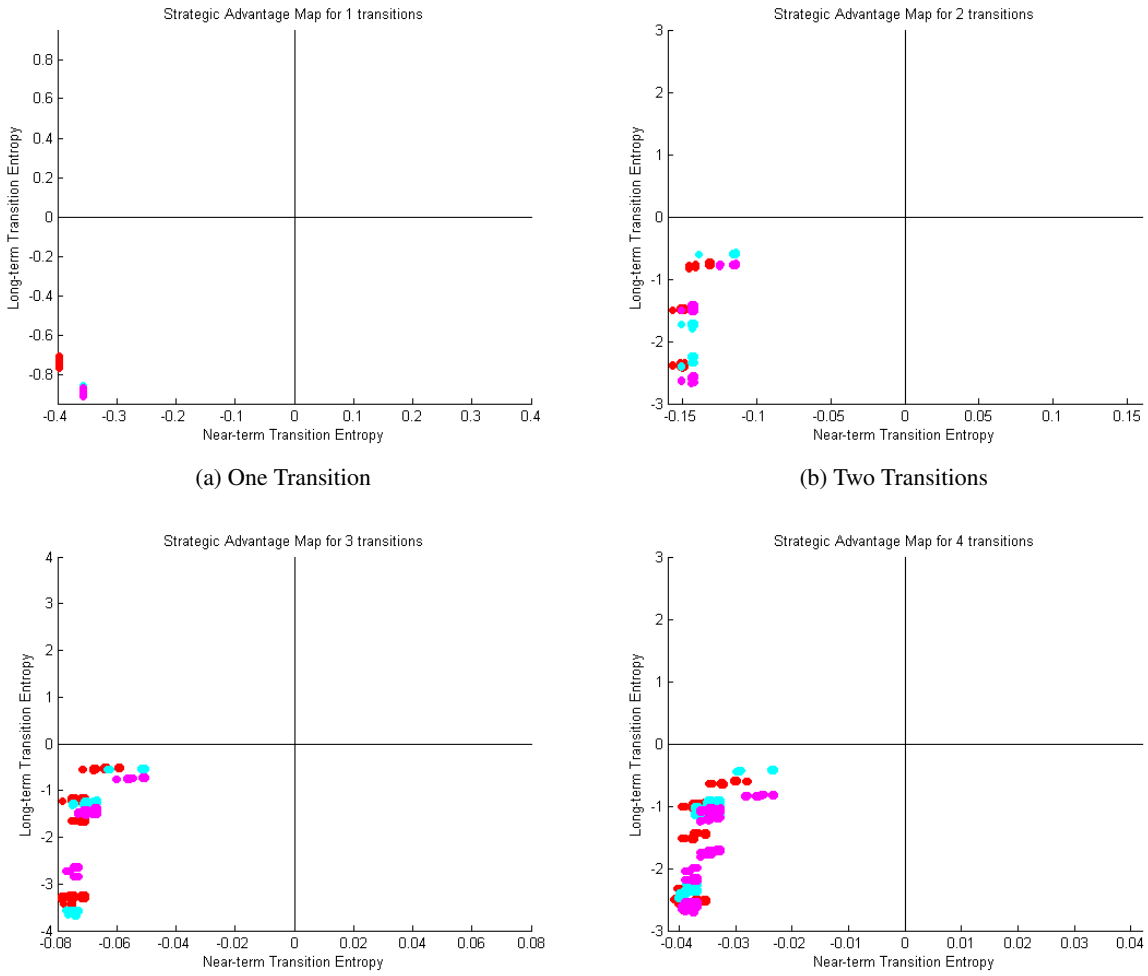
Figure 5.10: Strategic Advantage Map for Stage 3 looking forward four stages.

Decision for Stage 2: add Plane 26 to the network.

Stage 3

The results for Stage 3 are shown in Figure 5.10.

For all future stages, the option to add a LAN at the Smythii basin floor site (represented by blue dots) in this stage is either the most preferable or among the most preferable. Beyond the immediate transition, the add Plane 76 and add Plane 86 options begin to tag along with the add LAN option. It would appear that adding one of these two planes is likely to be the preferable option within another stage or two. Looking at the data in the access Table 5.5 would appear to confirm this hunch.



(a) One Transition

(b) Two Transitions

(c) Three Transitions

(d) Four Transitions

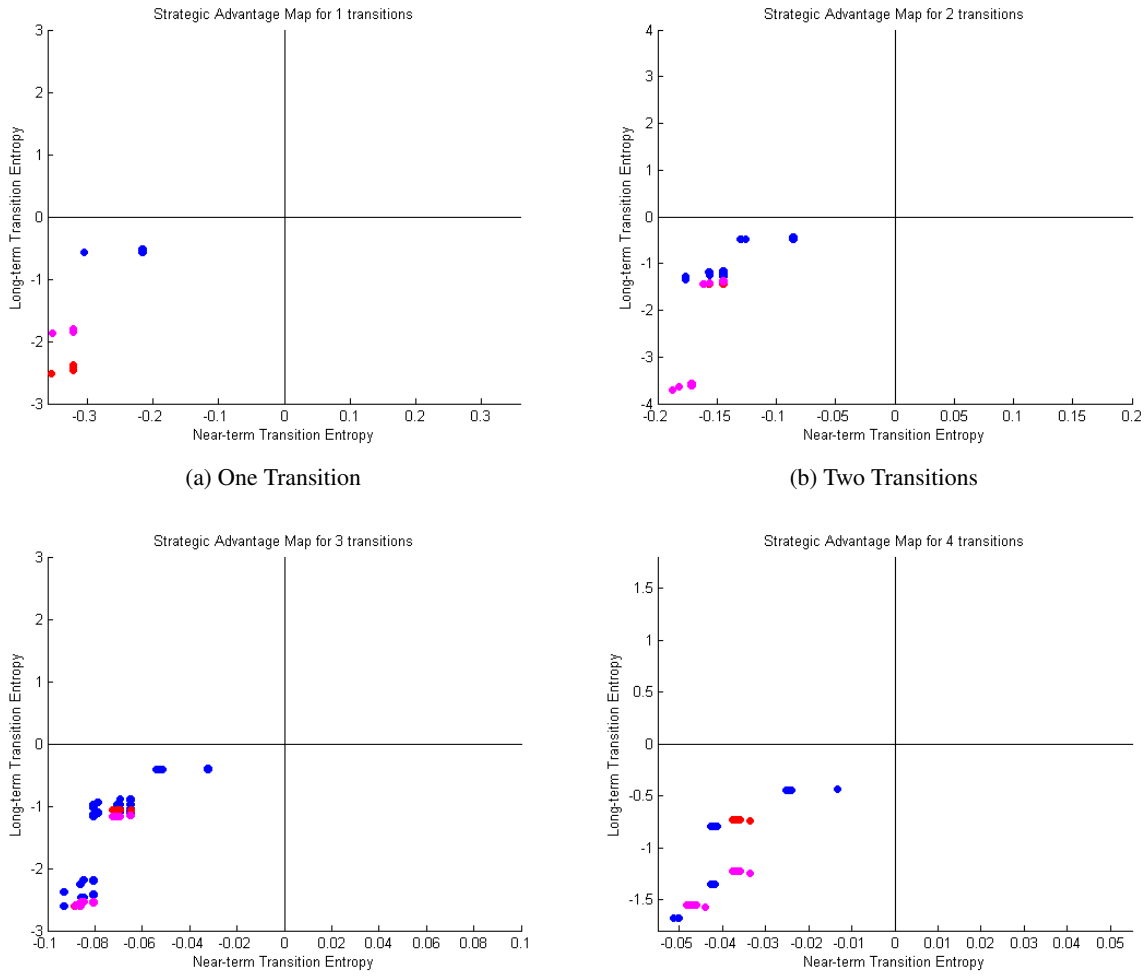
Figure 5.11: Strategic Advantage Map for Stage 4 looking forward four stages.

Decision for Stage 4: add a LAN at the Smythii basin floor site, and begin preparations to deploy either Plane 76 or Plane 86.

Stage 4

The results for Stage 4 are shown in Figure 5.11.

These maps indicate that adding Plane 76 (cyan) is now the optimal decision (based on transitions two through four). The first transition map is a little confounded by the add Plane 50 option. The immediate result is nowhere near as good as for Plane 76 or Plane 86, reducing its relative near-term entropy, but then it has several very good access values relative to either Plane 76 or Plane



(a) One Transition

(b) Two Transitions

(c) Three Transitions

(d) Four Transitions

Figure 5.12: Strategic Advantage Map for Stage 5 looking forward four stages.

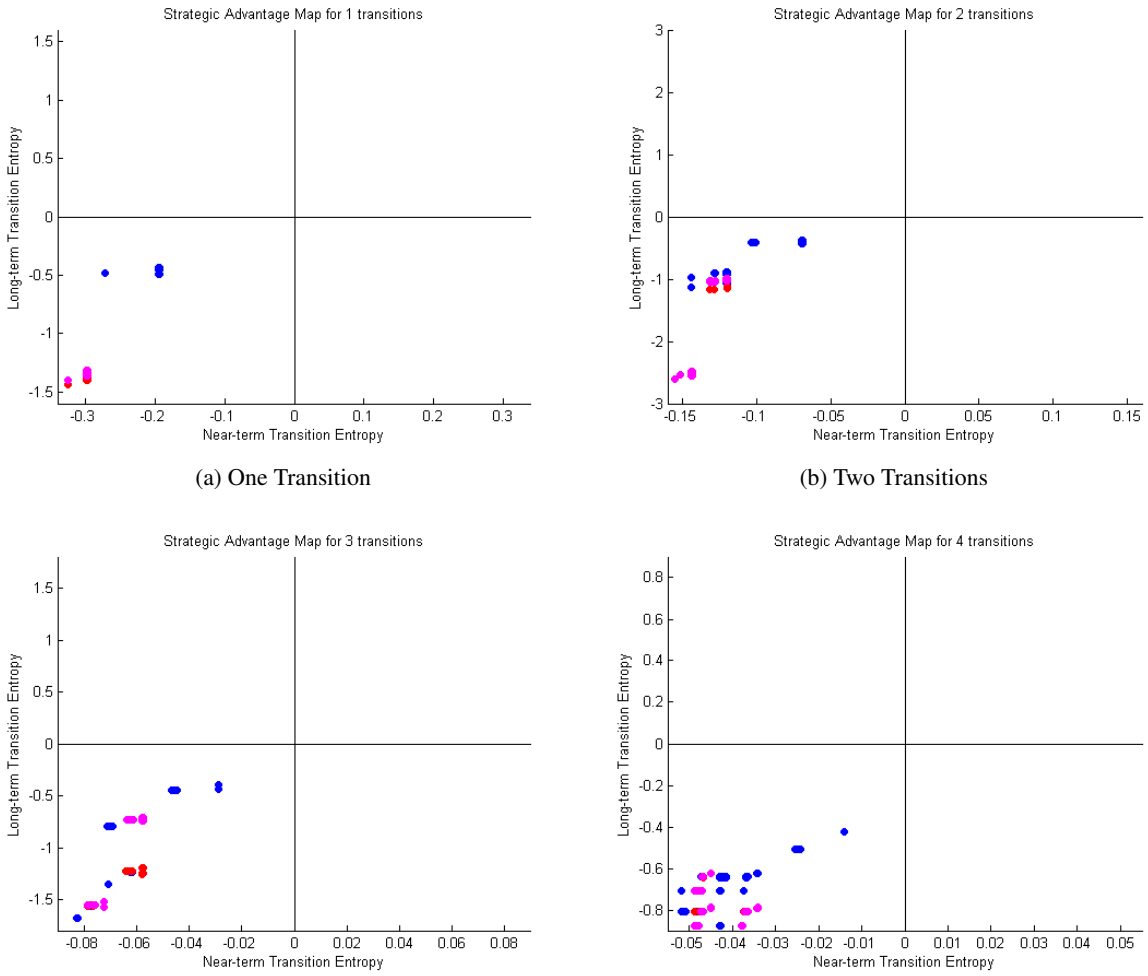
86, giving it a slightly better long-term position. Even in the other maps, it does relatively well, but nowhere near as well as the add Plane 76 option for meeting the needs of the central far-side highlands location.

Decision for Stage 4: add Plane 76 to the network.

Stage 5

The results for Stage 5 are shown in Figure 5.12.

The best option is to add a LAN to the South Pole since the blue dots representing the LAN only deployment dominate the figure.



(a) One Transition

(b) Two Transitions

(c) Three Transitions

(d) Four Transitions

Figure 5.13: Strategic Advantage Map for Stage 6 looking forward four stages.

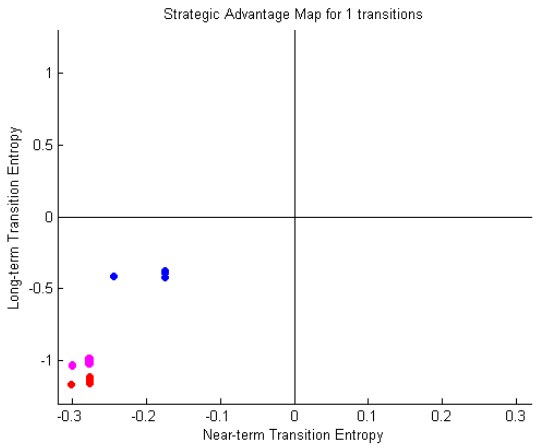
Decision for Stage 5: add a LAN to the South Pole site.

Stage 6

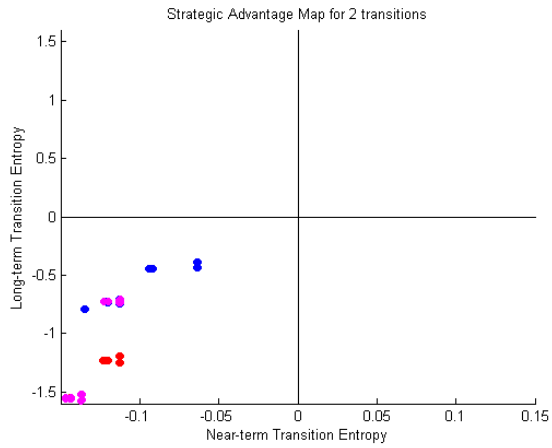
The results for Stage 6 are shown in Figure 5.13.

Again, the best option is to add a LAN to the selected location. It appears that the minimal constellation given the set of options considered in this example has been found.

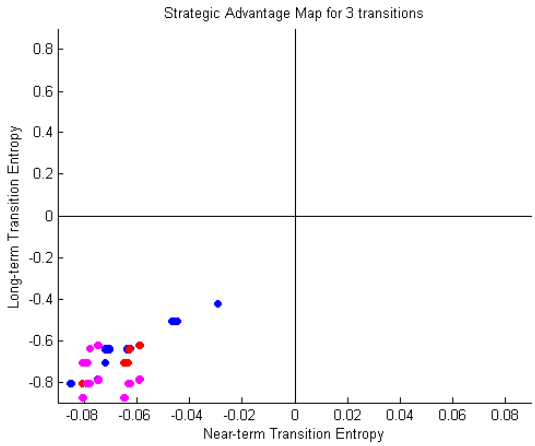
Decision for Stage 6: add a LAN to the selected location.



(a) One Transition



(b) Two Transitions



(c) Three Transitions

Figure 5.14: Strategic Advantage Map for Stage 7 looking forward 3 stages.

Stage 7

The results for Stage 7 are shown in Figure 5.14.

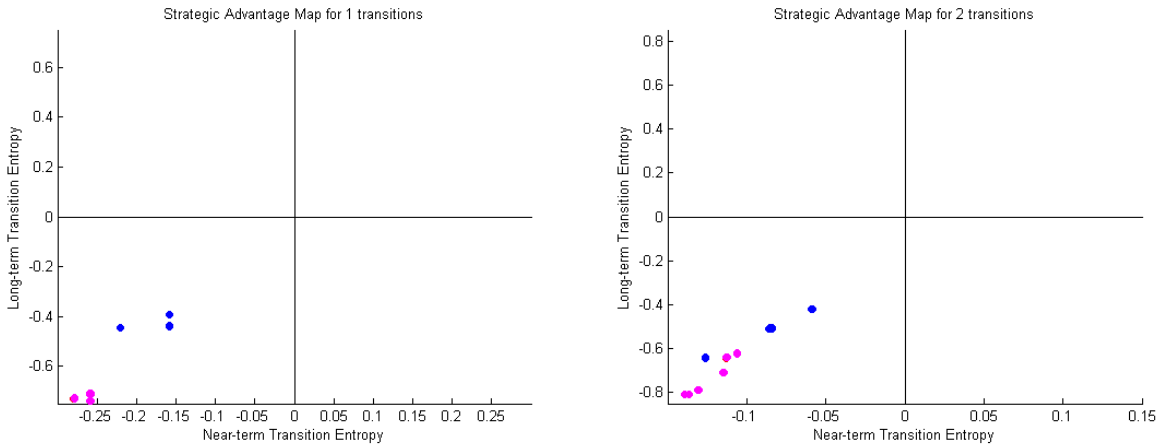
Again, the best option is to add a LAN to the selected location.

Decision for Stage 7: add a LAN to the selected location.

Stage 8

The results for Stage 8 are shown in Figure 5.15.

The best option is to continue to add LANs to each location visited in the remaining stages.



(a) One Transition (b) Two Transitions

Figure 5.15: Strategic Advantage Map for Stage 8 looking forward 2 stages.

Decision for Stage 8: add a LAN to the selected location.

5.4 Discussion

The minimal staged deployment communication architecture evolution given the considered set of network augmentations is LAN-LAN-(Plane 26)-LAN-(Plane 76)-LAN-LAN-LAN-LAN-LAN. In every stage, a LAN is deployed.

The continual position of these options in Quadrant III on the Strategic Advantage Maps indicates that the upfront deployment of the comparison constellation is the preferable option. The options considered here never performed well enough in comparison to move out of Quadrant III.

There are several obvious limitations to the results. The first limitation is the design of the constellation planes. These should be as optimized as much as possible, and then put through the strategic advantage methodology. The second limitation is the very rough estimation of the mass penalties and the simplistic benefit metric. Finally, there needs to be significant refinement of the relay estimations which in their current state are a limitation. The current simplistic assumptions tend toward worst-case results and may not be a realistic estimation of the true penalties and benefits. These limitations bias the framework against staged deployment.

Chapter 6

Conclusions

This chapter briefly summarizes the objectives, approach, major results and contributions of this thesis, evaluates its shortcomings, advantages and potential applications, and discusses potential avenues for future work and other recommendations for extending the research and case studies.

6.1 Summary

First, the objectives, approach, major results and contributions of this thesis are summarized.

The primary objective of this thesis is to develop a framework that directly addresses legacy challenges during system operation within the context of space communication networks. The first challenge is that legacy influences what changes are feasible and available as well as when a change may be made and how long it may take to implement. The second challenge is that legacy influences the effectiveness and desirability of any change that is made. Often this influence is a function of time and uncertainty. Finally, legacy influences the future constraints placed on the system as the result of making further changes.

In addition, the thesis aims to:

- Ascertain a set of principles based on historical examples that can guide decision making.
- Find a method to value a physical non-market traded legacy system.
- Find a method to value non-monetary legacy systems.
- Develop a tool to visualize the system's exposure to uncertainty if a transition is exercised.

In order to meet the desired objectives, the following approach was undertaken.

- **Observe:** historical and current systems (the NASA Deep Space Network (DSN) and the transition from IPv4 to IPv6).
- **Identify:** a set of principles from empirical patterns and insights.
- **Create:** a framework, methods and tools.
- **Apply:** the knowledge gained to global commercial satellite communication systems and the deployable communication networks to support manned lunar exploration.

The application of the framework to the two forward-looking case studies produce a few interesting results.

First, the Strategic Evolution of Systems framework might enable attainment of the optimal path over time as demonstrated in the global commercial satellite communication system case study. There are natural limitations to the performance of the framework in this regard, including the fidelity and accuracy of the underlying models, accurate accounting of the driving sources of uncertainty, the actual resolution of uncertainty as time progresses and the decision maker's skill in reading the advantage maps.

Second, the satellite communication system case study demonstrates that trends in the resolution of uncertainty — here, subscriber demand — are reflected in the strategic advantage maps. The position entropy formulation used to create the strategic advantage maps is based on forecasts of the uncertainty.

Third, the position entropy formulation is shown to be quite sensitive to the choice of debiasing factors. This result is discussed in more detail in the future work section.

Finally, the lunar case study applies the framework to systems in which the benefits and costs are not monetary in nature by establishing a benchmark architecture. The meaning of unit energy is redefined as the point at which the staged deployment architecture is as desirable as the benchmark of launching a full lunar constellation upfront with a pre-deployed ground network. This redefinition applies to systems that do not have the option to maintain the current system. The lunar networks are augmenting, never visiting the same landing sites twice.

The four main contributions of this research are:

1. Establishment of a new framework for the evolution of complex technological systems that naturally incorporates legacy.
2. Articulation of two important principles of system evolution based on empirical observation of NASAs Deep Space Network (DSN) and the transition from IPv4 to IPv6.
3. Preliminary formulation of a position entropy metric, derived from information entropy, to describe the current position and evaluate the desirability of potential future transitions of system configurations.
4. Application and implementation of the framework to two forward-looking case studies: a commercial satellite communication system, and the staged deployment of communication infrastructure to support manned lunar exploration.

6.2 Evaluation

This section evaluates the shortcomings, advantages and potential applications of the Strategic Evolution of Systems framework, methods and tools.

This thesis successfully addresses the legacy challenges by proposing a novel framework — perspective, position, plan and pattern — which is based on Mintzberg's emergent interpretation of strategy. The identification of feasible and available changes, when changes may be made and how long changes may take to implement are addressed in perspective. The architecture instantiation network provides a map of the feasible and available changes as enumerated using Willard Simmon's Architecture Decision Graph methodology. Specification of the evolution rules captures the time-related constraints of the change options.

The effectiveness and desirability of any change that is made and the impact legacy has on future constraints are captured in the concept of position. By evaluating the behavior of each possible configuration as a function of time, uncertainty and the path taken through the option space, it is possible to gauge the effectiveness and desirability — or the dynamic multi-dimensional value — attached to exercising each transition. By addressing the path-dependency challenge inherent to complex systems, it is possible to gain a sense of the future constraints imposed by legacy. Evaluating the possible paths looking forward indicates what the legacy system and any immediate change to it may enable or exclude downstream. The preliminary position entropy metric captures the dynamic multi-dimensional value attached to exercising each immediate option. The system's

exposure to uncertainty may be visualized by mapping the near-term and long-term advantage of each immediate option for each experiment. Here, advantage is defined as the difference between the position entropy of the option if exercised and the corresponding position entropy of the legacy system. The experiments are defined as some combination of the sources of uncertainty forecasted over the time horizon.

The theory of the Strategic Evolution of Systems was first developed by studying the historic evolution of the Deep Space Network using pattern-identification and insight by example. Two important principles for evolving complex systems with legacy were identified in the case study and further confirmed with additional empirical examples.

There are a number of similarities between the Strategic Evolution of Systems and Time-Expanded Decision Networks (TDN), but there are several distinct differences. As mentioned in the introductory chapter, the methodology is demonstrated using Heavy Lift Launch Vehicles, which are capable of launching more than 30,000 pounds to low-Earth orbit. In a sense, the TDN methodology models the switching costs for product platforms — or enables evaluation of whether a set of architectural options can be considered product platforms — but does not extend easily to capturing system-of-systems such as distributed modular networks. Rather, these vehicles are single integrated systems. Furthermore, the method is geared toward designing more flexible complex systems by identifying ways to reduce switching costs. Time-Expanded Decision Networks uses very similar steps as the Strategic Evolution of Systems, although the methodology does not specifically enumerate transition rules other than to say that the switches between configurations must account for the costs of all possible switches. Furthermore, it is assumed that the family or set of designs has already been found, whereas the Strategic Evolution of Systems provides steps for identifying these designs. The TDN methodology is intended for use in aiding in the initial design to improve the future evolvability of a system, and it focuses on minimizing life cycle costs.

The field of real options says that decision makers have the "right, but not the obligation," to exercise an option. The Strategic Evolution of Systems says essentially the same thing. The important underlying assumption of Real Options is that the source of uncertainty is a market-traded asset. Uncertainties inherent in complex technological systems rarely meet this condition. Kalligeros[87] notes the existence of multiple arguments to circumvent this assumption so that the well-established techniques in real options apply. The Strategic Evolution of Systems does not have this restriction: in a sense, the options are valued relative to the legacy or to the benchmark architecture. Kalligeros

also notes several other aspects of applying real options analysis to real systems that are not a limitation of the Strategic Evolution of Systems:

- "The uncertainty and recourse actions (i.e., system re-configuration) are all of a similar time scale." (In Strategic Evolution, the time scales can be established according to the needs and nature of the system and its evolution.)
- "The system is designed according to a 'platform' architecture, on which flexible components can be added, removed or exchanged at finite cost." (This is similar in concept to Time-Expanded Decision Networks, where the system acts like a "platform" and is more integrated than truly modular. The Strategic Evolution of Systems can, in theory, evaluate platform architectures, but it can more generally evaluate distributed system-of-systems. The framework is not limited to "plaftform" architectures or single integrated systems.)
- "There is no ambiguity as to the appropriate partition of the design vector into $x_t = x_t^B; x_t^F$. As mentioned above, optimizing over the selection of the 'platform' and the re-configurable parts of the architecture becomes a very large problem." (Strategic Evolution extends this by identifying those aspects of the design (decision) vector that are irreversibly changeable. The base design vector is similar to the idea of those aspects that are not changeable unless a different architecture family is used. The flexible design vector acts as the fully reversible design decisions. The real options framework as described here does not account for irreversible behavior.)
- "The uncertainties are market-traded and the systems' cash flows are determined by market equilibrium." (Again, the Strategic Evolution of Systems does not have this restriction due to the definition of unit energy and the subsequent conversion to the entropy transform for insight. Although the Strategic Evolution of Systems could be quantitative, it focuses more on qualitatively identifying behavior to value the options relative to the legacy or the benchmark architecture.)

One assumption of the Real Options framework proposed by Kalligeros that is similar to the Strategic Evolution of Systems is the following: "The number of alternative re-configurations is relatively small and known in advance in terms of the underlying uncertainties and configurations." The Strategic Evolution of Systems framework may extend this to some degree since the option tree, objectives and uncertainty could potentially evolve dynamically.

There is a very strong connection between the Strategic Evolution of Systems and the fields of decision theory and decision analysis.[88] The goal of decision theory is to find the optimal decision. Decision analysis is concerned with providing the tools to guide decision makers to the best decision possible. There are numerous aspects of decision theory: making decisions under uncertain conditions (e.g., expected value, Bayes theory, subjective probability and etc.), the temporal relationships between decisions and their outcome (e.g, inflation, discount rate), the behavior and relationship between "competing decision makers" (e.g, game theory and signal detection theory) and decisions that must be made in complex situations or for complex systems. Many decision analysis tools make use of statistical analysis, including evaluation of risk. The power of Strategic Evolution of Systems is that it incorporates many of these aspects in a way that naturally includes legacy. An interesting extension of the Strategic Evolution of Systems framework could examine whether the influence of competing decision makers can be incorporated into the framework.

Finally, there is an interesting connection between the Strategic Evolution of Systems and stakeholder network mapping. According to Svendsen and Laberge:[89]

> *A stakeholder network map is a tool for identifying who has an interest or "stake" in an issue. Network maps can also be used to show the quality of relationships that exists between stakeholders.*
>
> *Mapping helps us to see the whole system around an issue by making the network of existing relationships visible. It can also help us gain a better understanding of the quality of those relationships and the leverage points for building or strengthening the network over time.*

The stakeholder network mapping seems similar in concept to the architecture instantiation networks and the identification of supernodes. Instead of examining the relationships between architectural configurations, stakeholder networks map the relationships and behavior of social networks. Stakeholder networks have been used to isolate potential areas of conflict so that decision makers can formulate strategies to mitigate the delays and other risks associated with discord among people.

6.2.1 Shortcomings

This section describes five inherent shortcomings to the framework, methods and tools as presented in this thesis.

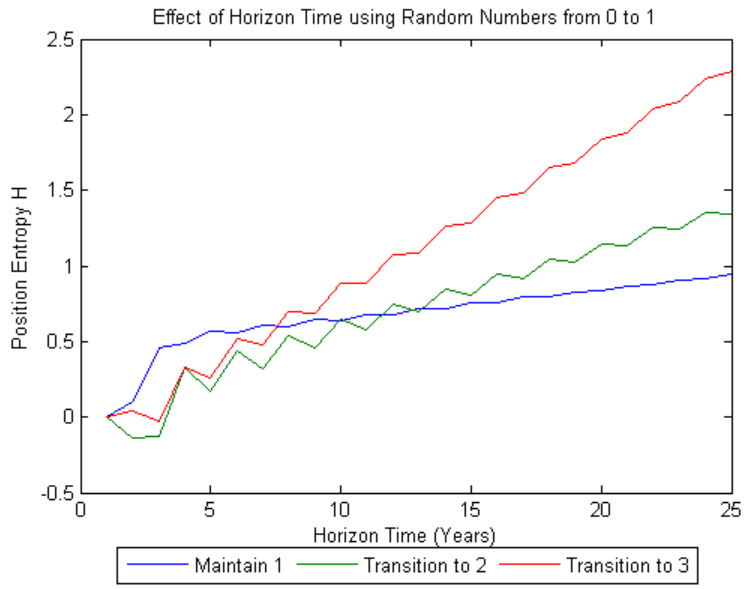


Figure 6.1: A plot of ending position entropies over a range of horizon times for the case in which all variables are randomly chosen on a range of zero to one.

The primary disadvantage is the computational complexity required to evaluate the large number of option paths for each experiment — and there may be a large number of experiments depending on how many sources of uncertainty there are, the number of time steps to consider and the desired reliability of the forecasts. The preliminary entropy formulation requires propagation of the energy along every option path for each experiment out to the time horizon, calculation of the terminal entropy values and a subsequent propagation of the entropy metric from the horizon time back toward the current time. This analysis requires a significant number of calculations, scaling quickly with the number of options. Since the performance of the system configuration is dependent on so many things — time, uncertainty and path through the option space — the model fidelity and complexity is a corresponding issue. To paint an accurate picture of the dynamic behavior of the system and its options, the underlying models become intricate to construct.

In the course of developing the Strategic Evolution of Systems, it was found that the preliminary position entropy formulation is highly sensitive to the choice of horizon time. This sensitivity is highlighted in Figure 6.1. The relative ranking of the available options — based on their position entropies — switches several times as the horizon time is increased.

The significant computational burden imposed by the formulation required a method to make the

problem more tractable in simulation. The solution was to artificially scope the set of options, narrowing them down to a limited set that, by inspection, showed the most promise. Since the number of options was artificially limited, the number of option paths associated with each available transition was also artificially scoped. This caused the position entropy formulation to bias towards configurations that had a large number of outgoing options and away from configurations that appeared to act as sinks due to the scoping but in reality had several options that were simply not considered. This problem arose because the entropy formulation acts as a superposition and the more paths that are evaluated, the more weight the immediate option tends to receive. The effect was a kind of double counting of the flexibility as measured by the number of available options out of a given configuration. Thus, it was necessary to create a debiasing factor that would reduce this artificial double-counting effect. Similarly, the configuration — flexibility — term is a function of the cardinality of the set of options, a unit fraction. The other terms could be more or less or equal to unity and are not limited to unit fractions. Unit fractions tend to grow more strongly in a logarithm than their counterparts in the other terms. For example, $-log(1/5) = log(5)$, which is not insignificant. As the number of options grows, this term grows quickly and can easily drown out the contributions from the other terms for networks with many options. To compensate, a configuration debiasing factor was introduced to reduce the impact of the unit fraction and to keep it at roughly the same level of contribution as the other terms. In the course of using the preliminary entropy formulation with these debiasing factors, it was found that the results were highly sensitive to them.

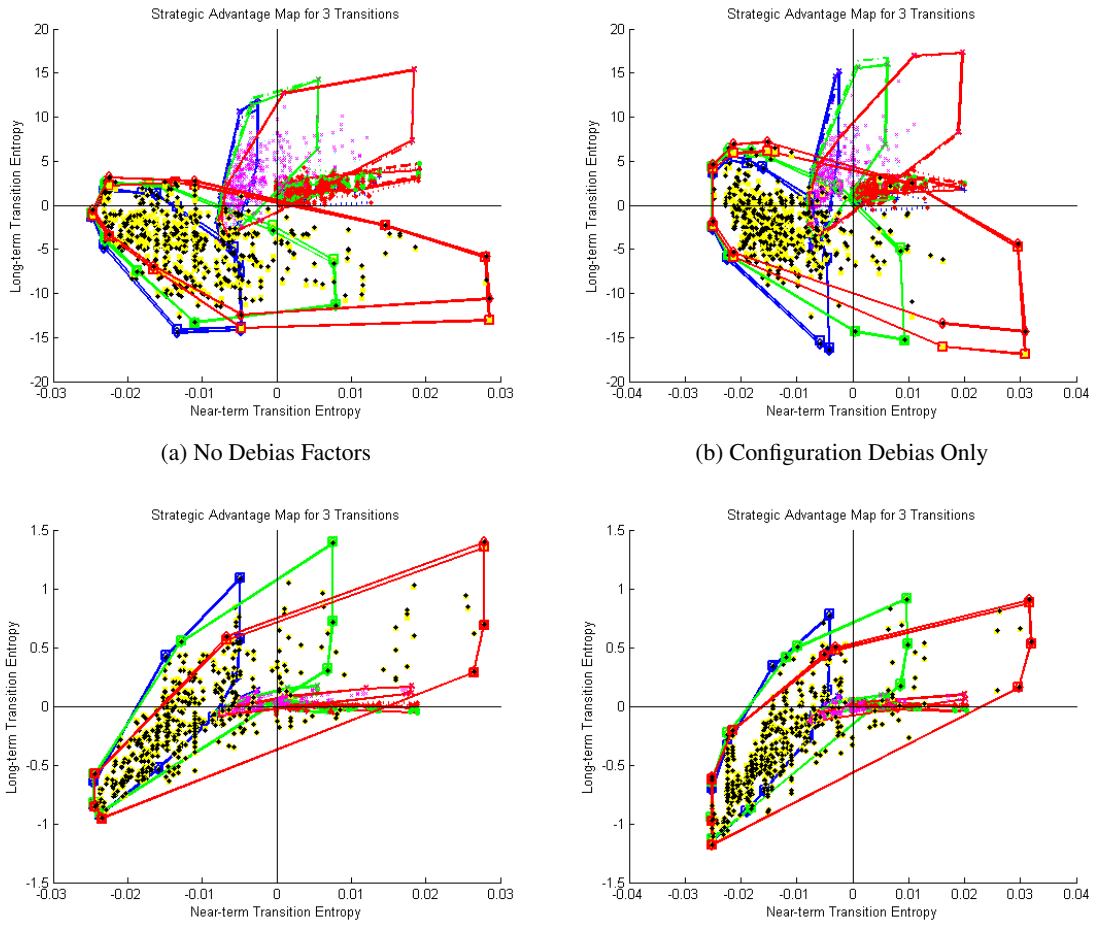
To recap from Chapter 4, Figure 6.2 shows the effect of the configuration debiasing factor and the number of option paths debiasing factor separately and combined. The example is for the first time period (the initial Strategic Advantage Maps) looking out three transitions. The three architectures comprising Constellation 1 are blue, green, and red; the two architectures comprising Constellation 2 are cyan and magenta, and the two architectures comprising Constellation 3 are yellow and black.

The most noticeable difference is the significant change in shape and scale when the number of option paths is scaled according to the debiasing factor described in the Weightings section. The long-term entropy axis shrinks from 20 to 1.5. The near-term axis is consistent because of the way the near-term entropy is estimated. Since the number of options available to the lower constellations are artifically scoped — seven options when in Constellation 1 reduces to four options in Constellation 2 and further reduces to two options in Constellation 3 — the number of option paths are also artificially scoped and this can bias the entropy formulation in favor of remaining in Constellation

1 when this may be a poor choice relative to the other constellations.

The configuration term is not nearly as significant though it does appear to contract the graph slightly toward the origin along the long-term axis. Without the option paths debiasing factor, the configuration term separates the regions specificed by the different (immediate transition) architectures in Constellation 1. With the option paths debiasing factor, incorporating the configuration debiasing factor appears to level out the long-term entropies along the top of the graph.

The fourth shortcoming involves limitations with visualizing the system's exposure to uncertainty. The strategic advantage maps can quickly become overwhelming with detail even for systems with



(a) No Debias Factors

(b) Configuration Debias Only

(c) Number of Option Paths Debias Only

(d) Both Debias Factors

Figure 6.2: A demonstration of the effect of the debiasing factors used in this case study. The figures represent Time Period 1 looking out three transitions with (a) no debiasing factors, (b) the configuration debiasing factor only, (c) the debiasing factor for the number of option paths only, and (d) both debiasing factors considered simultaneously.

only seven configurations in consideration. Filtering by the immediate transition exercised helps to some extent, but this method creates a lot of subgraphs to examine and analyze.

Finally, in the course of propagating the methodology through time — in simulation — it was found that the Strategic Advantage Maps require detailed understanding of the assumptions, evolution rules and models in order to appreciate the meaning of the relative behaviors of the system and their trends over several near-term transition exposures. This appreciation is necessary to identify the dominating immediate transitions.

These shortcomings will be discussed further in the future work section.

6.2.2 Advantages

This section describes five advantages of using the framework and tools presented in this thesis.

First, the framework naturally incorporates legacy and the path-dependency inherent in complex technological systems. The identification of the origin in the Strategic Advantage Maps with the legacy system and the philosophy of measuring position relative to the legacy is a recognition of the idea of legacy as a "hindrance," or the tendency of decision makers to maintain a system "as is." A transition option must be strongly dominating over the legacy for the decision maker to choose to exercise it. The value of legacy is found by evaluating its desirability, or its dynamic multi-dimensional value over time, uncertainty and the path taken through the option space. Identifying what the legacy enables and excludes is based on the structure of the option network includings its supernodes and the relationship this structure has with the system desirability.

Second, an important benefit of the Strategic Evolution of Systems framework is the ability to identify the most advantageous time to exercise options by way of the Strategic Advantage Maps. These maps plot the relative behavior of the immediate transition options as a function of the uncertainty, generating regions of possibilities that can be compared in terms of their range of relative desirability. Examining these maps over a range of near-term transition periods (e.g., one year out, two years out, and three years out) enables the emergence of trends that can be used to gauge the appropriate allocation of resources in preparation for the implementation of a transition option in the near future. The degree of dominance on the map and the trends and shifts in dominance in the future can help determine this allocation. When it is clear that a transition option is dominating, then the decision makers know that it is time to exercise that option.

Third, the application of the preliminary position entropy formulation enables inclusion of multiple sources (dimensions) of uncertainty. A design of experiments could be run over these multiple sources in order to gain insight into the behavior of the configurations by sampling the space of possible outcomes in an intelligent and efficient manner.

There is an inherent time/uncertainty/path-dependency coupling. The further out in time that is forecast, the greater the number of possible option paths and the wider the cone of uncertainty. The energy of a system is a function of the assumed resolution of uncertainty — as predicted by a given experiment — and the option path. As the energy is propagated through time, the variables representing the "resolved" uncertainty and the option path are continually changing. The way in which these variables are accounted for enables incorporation of time-dependent uncertainties and options. For example, configurations could become obsolete (option disappears) in three time periods and a new technology could emerge (option appears) in four time periods.

Finally, the Strategic Evolution of Systems framework seeks to gain overall insight into the relative behavior of system configurations rather than focusing on a single, possibly misleading metric. By looking at high-level structure and behavior in the form of the system's exposure to uncertainty diagrammed in the Strategic Advantage Maps, it is possible to jointly evaluate the impact of multiple, coupled and intricate aspects of the system and to do so along multiple dimensions.

These advantages will be addressed further in the future work section.

6.2.3 Applications

Although this thesis focused on applications to space communication networks, it may be applied to myriad other systems.

The theory and framework described in this thesis was developed for communication infrastructures which may be described as distributed modular system-of-systems, but, in theory, it may be extended to any system that is created in terms of discrete objects or processes. The continuous nature of roadways or walls in a building is misleading. To construct such apparently continuous structures, discrete processes are followed. To make a wall, the act of installing drywall pieces using screws into the wall framing is followed by plaster application, sanding and painting. The theory described here can be used for discrete assets or discrete processes. The theory is essentially a description of taking the pieces of the puzzle, identifying the structure and observing the patterns in behavior that appear over time in order to make more informed decisions.

Space communication networks have a natural limitation in that the hardware is inaccessible. Thus, the set of transitions was assumed to be limited to reconfiguring the existing assets, augmenting the existing assets with additional assets or making software adjustments. These limitations are usually not an issue for terrestrial systems. This assumption does not necessary apply to the framework, methods and tools, and will be discussed in more detail in the future work section.

Generally, the framework is meant to provide insights for complex systems, in which the best transition is not obvious. For simple systems, it may be possible to analyze the instantiation network and get a clear sense of the best option by inspection. Large and complex systems may require the considerable and intricate analysis required to use this framework.

Perhaps more significantly, this framework is meant to be used for systems in which it is necessary to gauge the best time to make a change — pre-determined or otherwise.

6.3 Future Work

This section discusses potential avenues for future work to extend and refine the research presented in this thesis.

A significant number of the systems that would benefit from this framework are complex problems with large and complex option networks where the best decision is not obvious by inspection. However, these systems suffer from the computational requirements of the framework as presented in this thesis. It will therefore be necessary to investigate numerical techniques to improve the computational efficiency. One possible solution is to apply a Viterbi decoding to intelligently search the "messy bush" for the non-dominated subset of paths instead of a full search of all paths (Viterbi **does not** throw away dominated paths). There are several possible uses for a technique such as this: first, it could be used to intelligently focus the number of options to a non-dominated set; second, given a reduced set of options, it could be used to extend the method out to 50 or more steps rather than the 10 considered in this thesis; third, identifying the non-dominated paths could potentially be used to simplify the Strategic Advantage Maps.

The preliminary position entropy formulation for measuring the relative near-term and long-term positions of architectural options is unique. This thesis has only scratched the surface of its mathematical complexities and implications. Significant work remains to determine energy functions

appropriate to the system in question, as well as the development of methods to debias and calibrate the formulation. The alpha term reduces the impact of the integer cardinality, the gamma term reduces the impact of the number of option paths and the beta terms are decision-maker defined weightings. It will be necessary to study the effect of varying the beta weightings and scaling on the outcome and to identify methods to calibrate the beta terms using the preferences of the decision makers. One solution is to use utility theory to identify appropriate utility curves to sample as the situation evolves. An important question is how to incorporate the discount rate. There appear to be two options: Beta 4 — the future term weighting — can be adjusted as a function of the time period in order to account for the discount rate anticipated for that period, or the discount rate can be explicitly incorporated into the energy accounting as the costs and benefits are accrued over time.

The preliminary position entropy formulation has been shown to be sensitive to the chosen time horizon. For this thesis, the horizon time was set to the system lifetime, even though it is a separate issue altogether. It is unclear how to choose an appropriate time horizon. Thus, it will be necessary to identify a method to consistently choose the best time horizon. Are there factors that shape the optimal time horizon? What are reasonable justifications for choosing the horizon time? In the course of the research, it was hoped that a settling effect over time would be observed, but this was not the case. Is it possible to design the terminal entropies, energy equations or even a reformulation of position entropy to shape the curve such that there is a settled horizon time?

In this thesis, it has been assumed that the only uncertainties are mission-level (e.g., landing sites) or economic (e.g., demand level). Other sources of uncertainty will need to be considered, such as the reliability of assets and the risk associated with using new technology. These additional sources of uncertainty can be incorporated using a multi-dimensional experimental design. The framework could therefore accommodate these multiple sources by sampling the space of possible outcomes in an intelligent and efficient manner (e.g., latin square block design, fractional factorial design and etc.) in order to gain insight into the behavior of the configurations with as few experiments as possible.

A promising area of future work would extend the framework to allow dynamic reconfiguration of the option tree. This reconfiguration would support two aspects: adding options (leaf nodes or branches) mid-stream that did not exist in the first time period (t = 0) — for example, to model technology infusion or anticipated shifts in policy — and eliminating options (leaf nodes or branches) mid-stream to model obsolescence of technology or component and asset failures. These additions

and subtractions could be pre-planned or random. It should not be difficult to incorporate dynamic reconfiguration, though it will increase the computational load. In the case of random reconfiguration, the option tree can be reconfigured during the propagated time period that the random addition or subtraction occurs (it's unexpected). This could be achieved by simply adding or subtracting subsets of the option paths when time has propagated to the period in question. If the reconfiguration of the option tree is expected, then the set of options could be constructed upfront using an evolution rule to capture the expected addition or subtraction. A more interesting case is to observe the effect of planned or expected reconfiguration of the option tree occurring at an uncertain time. Here, it would make sense to make the structure of the option tree a dimension of uncertainty for the experiments, possibly creating a need for an additional filter for the Strategic Advantage Maps so this effect can be isolated and analyzed.

An important simplifying assumption used in the case studies is that the system elements are homogeneous. The framework provides a method to examine the impact of hybrid systems as demonstrated explicitly in the lunar case study in which a homogeneous set of ground assets can augment a separate homogeneous set of satellites assets. At the decision level incorporating heterogeneous elements and user devices could be done by specifying homogeneous subsets as an add/remove decision. The difficulty arises with modeling the effect that a varying heterogeneous system produces on the performance of the system as a whole.

Currently, the value of legacy is qualitative, based on a visual inspection of the change exposure visualization tools over several periods. However, future work could look at ways to quantitatively value the legacy based on these maps, perhaps as a weighted average of the transition entropies over a specified set of conditions. Systems such as the lunar case study that do not have a maintain "as is" comparative option would be able to quantitatively value the benchmark system using a similar measure. The visualization tools (Strategic Advantage Maps and the Iceberg Exposure diagram) are currently limited to about seven options using MATLAB. Including more options would increase the clutter in the Strategic Exposure Maps and Iceberg Exposures and make it even more difficult to see the patterns. Future work should investigate better ways of visualizing the position data, particularly if additional sources of uncertainty are modeled. There may be mathematical relationships within the outcomes that could provide qualitative support to the decision makers. At the moment, the manner of reading the maps and evaluating the meaning is completely qualitative. Although the maps are quite illustrative, it is challenging to fully understand the meaning of the patterns. This

will be an acquired skill, as it requires deep understanding of the nature of the transition rules, the evolution rules and the manner in which the position (e.g. entropy formulation) is calculated. It would be extremely useful to perform studies to more deeply understand the fundamental nature of the mappings so as to identify methods to quickly analyze the patterns without needing to be as well-versed in the underlying assumptions and rules.

For simplicity, the case studies in this thesis have assumed that the rip-out-and-build-new option is constrained to be within current architecture family. However, the large expense of ripping out and building new would be more likely to motivate a change to a different architecture family that is perceived to be more desirable. It is unclear how to identify more desirable architecture families. A framework similar to Strategic Evolution of Systems could be used and possibly justified from the considerable expense and risk associated with ripping out and building new, but there should be easier, less complex methods to identify a more desirable architecture family and specific architecture to consider for a transition. It would be interesting to find efficient ways of making this evaluation. An approach may examine the architecture instantiation networks associated with each possible family and identify a preferred architecture family by inspection. Even so, which architecture is the most preferred? What is the best trade between future flexibility, benefits and resources expended to achieve those benefits and flexibility? A method similar to how the best build architecture is currently identified could be considered (e.g., Multidisciplinary System Design Optimization).

A very important piece of future work would be to extend the Strategic Evolution of Systems framework to systems other than space communication networks. There are two main directions that this research could take. If the system in question is a terrestrial distributed modular infrastructure there are a few key differences. First, terrestrial systems have the benefit of access to components, unlike its space counterpart where the hardware assets are extremely difficult to modify. This increase in access would have the effect of increasing the number of options available to consider, thereby increasing the expected computational burden. Second, there may be considerable integration with external systems (e.g., network effect). Identifying available transitions, estimating the cost to make the change and modeling the impact (benefit, cost, risk, opportunity) of the change will be considerably more complex as a result. Second, the system may be a single integrated system, rather than distributed and modular. In this case, the research may follow the Time-Expanded Decision Network methodology[15] and its breakdown of identifying switching costs between single integrated

systems.

The thesis currently assumes that the legacy system continues to operate while a given transition is implemented. However, there are situations when the legacy system must be taken off-line for some period of time during the transition. If the system must be kept off-line for the entire time period, then the energy accounting equations can be adjusted to account for the cost of transitioning and no benefit for keeping the system off-line. Risks could be modeled by looking at a range of possibilities for the costs during the transition period when identifying appropriate experiments. In some systems, the future benefits will be impacted as a result of taking the system off-line for a period of time (i.e., lost customers). The risks of this impact can be incorporated in a similar way. Other systems may only need to be off-line for a fraction of the time period. This time period could be broken into multiple, smaller time periods and a series of transitions could be used to model the differences between how each of these periods is treated.

It has been implicitly assumed in this thesis that the decision makers have a single objective. Multiple objectives could be jointly considered using methods from multidisciplinary system design optimization. Another solution that may enable time-varying objectives is to consider a Hilbert space, wherein each <x,y> inner product space represents the impact to a given objective. In this representation, the exposure to uncertainty is evaluated for each <x,y> inner product, resulting in a Strategic Advantage map similar to those presented in this thesis. The underlying models and experiments are identical for each objective; the <x,y> inner product merely filters the results to show the impact to a single objective. The dominance pattern of transitions for each objective could be used to sort the options based on decision maker preferences (e.g., using utility theory). A quantitative valuation of the legacy and its alternatives would be key for this analysis. A single-metric performance of each option as a function of the objective could be plotted on a radar plot to identify dominances over a range of objectives. It may be possible to apply fuzzy math to incorporate some multi-dimensional measure of the exposure (similar to the Strategic Advantage Maps) onto this radar plot.

6.4 Closing Remarks

This thesis has introduced the Strategic Evolution of Systems, a novel framework for evolving complex systems that directly addresses legacy challenges during system operation within the context

of space communication networks. The framework — perspective, position, plan and pattern — is based on Mintzberg's emergent interpretation of strategy.

The Strategic Evolution of Systems framework and principles were used to evaluate a system's current position as well as to update the evaluation as time progresses. The satellite communication case study provided one example where the methodology enables identification of the optimal transition path over the system's operational life. The lunar infrastructure deployment illustrates how to redefine the meaning of unit energy to evaluate non-monetary systems and/or systems that do not have the option to maintain "as-is."

Generally, the framework is meant to provide insights for complex systems in which the best transition is not obvious. For simple systems, it may be possible to analyze the instantiation network and get a clear sense of the best option by inspection. Large and complex systems may require considerable and intricate analysis to use this framework in any meaningful way.

Perhaps more significantly, this framework is meant to be used for systems in which it is necessary to gauge the best time to make a change — pre-determined or otherwise.

It is evident that the choice of horizon time and the use of debiasing factors can have a significant influence on the results. Future study on properly identifying and constructing these variables is strongly recommended.

Finally, the ideas and tools presented in this thesis may be used to compare preferred systems to suggested alternatives in order to justify expenditures or to initiate research and development programs.

Appendix A

Acronyms

NAT Network Address Translation 41

IP Internet Protocol 60

IPv4 Internet Protocol version 4 17

IPv6 Internet Protocol version 6 17

TCP/IP Transmission Control Protocol/Internet Protocol 41

EELV Extended Expendable Launch Vehicle 41

HLLV Heavy Lift Launch Vehicle 41

VHS Video Home System 42

U.S. United States 40

VOR Value-Opportunity-Risk 47

COTS Commercial Off The Shelf 59

MAC Media Access Control 60

QoS Quality of Service 58

IT Information Technology 62

DSN Deep Space Network 16

IGY International Geophysical Year 25

DoD Department of Defense 25

CSAGI Comite Speciale de l'Annee Geophysique Internationale 70

USNC U.S. National Committee 71

WWII World War II 71

U.S.S.R. Union of Soviet Socialist Republics 72

JPL Jet Propulsion Laboratory 25

WRE Weapons Research Establishment 88

NITR National Institute for Telecommunications Research 88

DSIF Deep Space Instrumentation Facility 16

TRACE Tracking and Communications Extraterrestrial Network 93

SFOF Space Flight Operations Facility 16

TDA Tracking and Data Acquisitions 16

OTDA Office of Tracking and Data Acquisitions 16

OSSA Office of Space Science and Applications 16

LPPO Lunar and Planetary Projects Office 16

GCF Ground Communications Facility 16

VMCC Viking Mission Control Center 16

MCCC Mission Control and Computing Center 25

NOCC Network Operations Control Center 16

GSC Ground-SpaceCraft

DSS Deep Space Station 319

Ap.U. Aperture Unit 324

dB Decibel 246

IBM International Business Machines Corporation 331

FPSO Flight Projects Support Office 331

SFOC Space Flight Operations Center 331

AMMOS Advanced Multi-Mission Operations System 331

LAN Local Area Network 118

GPS Global Positioning System 28

INCOSE International Council on Systems Engineering 126

SMAD Space Mission Analysis and Design 27

ADG Architecture Decision Graph 127

VSS Very Small Station 29

MSS Medium-sized Station 24

LSS Large-sized Station 29

MUSE Mission Utility Simulation Environment 264

STK Satellite Tool Kit 252

W Watt 259

UPV Unpressurized Vehicle 259

HPI Human nature, Politics, and Influence

Appendix B

The Deep Space Network

B.1 DSN Process Models

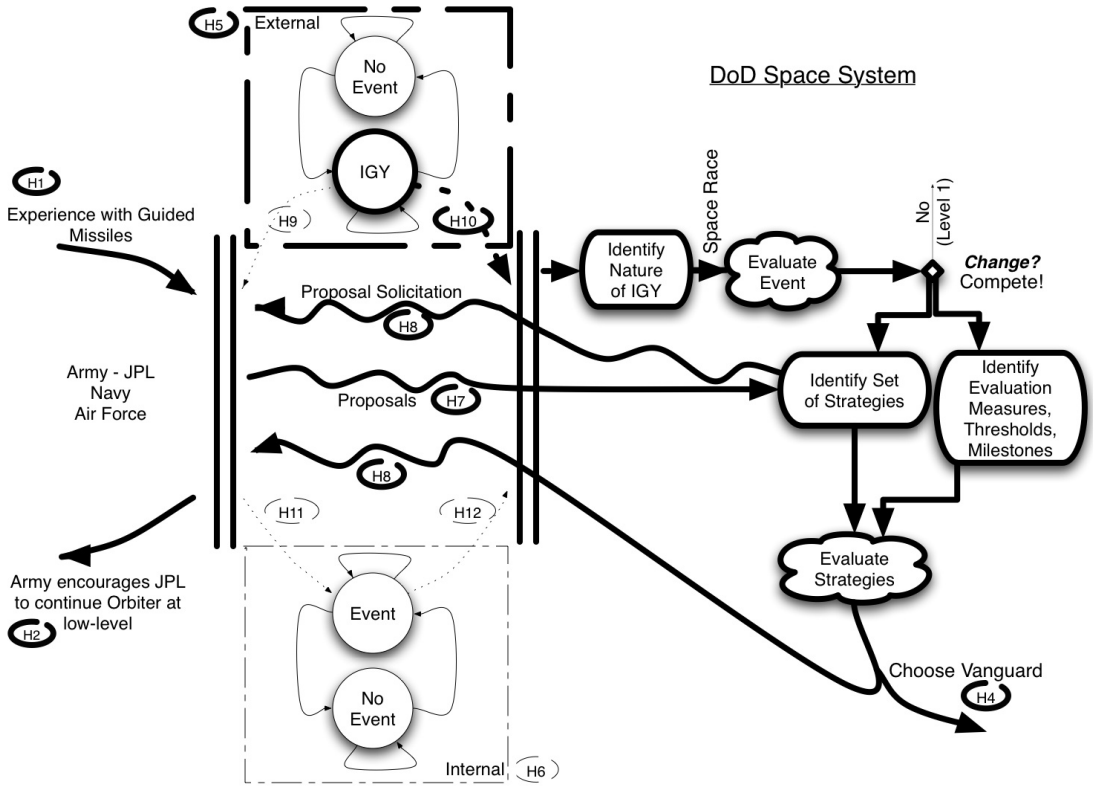


Figure B.1: Event flow for DSN Creation: The International Geophysical Year. For reference: International Geophysical Year (IGY), Department of Defense (DoD), Jet Propulsion Laboratory (JPL)

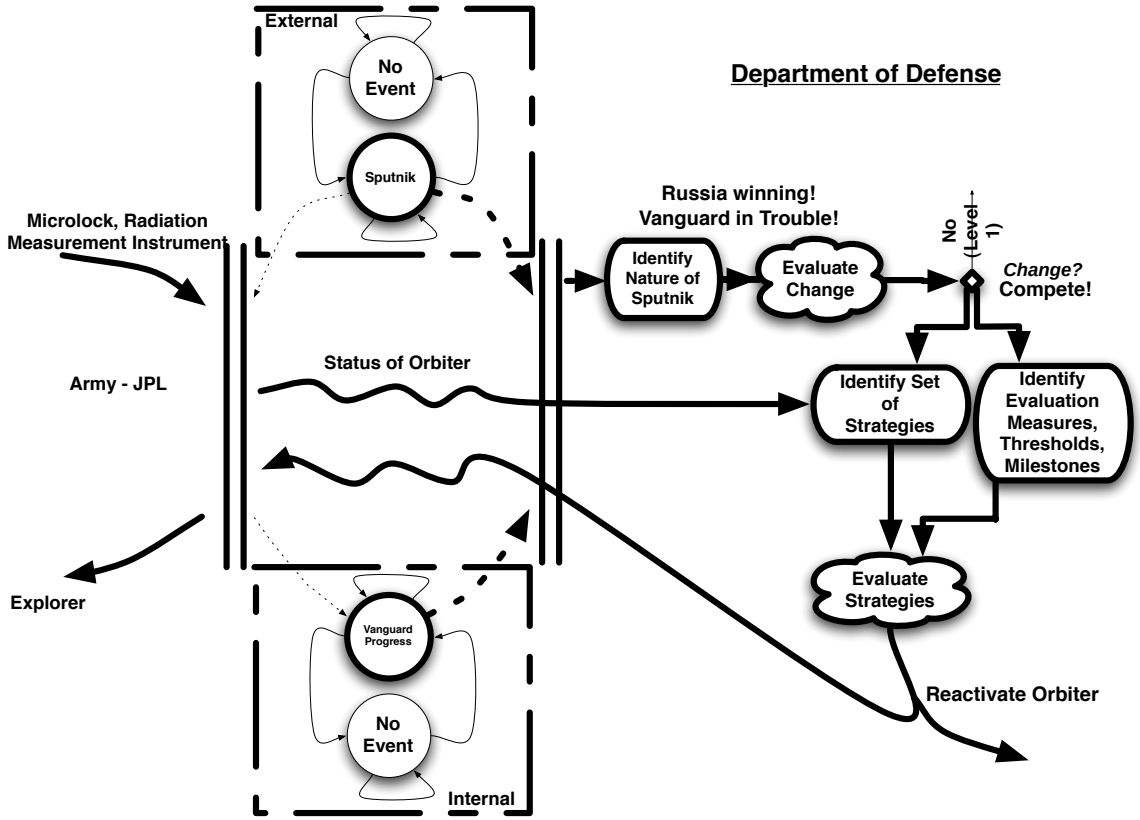


Figure B.2: Event flow for DSN Creation Stage 2: The Launch of Sputnik.

B.2 DSN Mission Timeline Tables

DATE	EVENT	NOTES	MISSION STAGE
1958 Jan. 31	Launch of Explorer 1.	Discovered Van Allen Radiation Belt w/ radiation instrument.	Earth P1B
1959 Mar. 4	Pioneer 4 moon fly-by.	Knocked off course by booster rocket malfunction. Camera failed. Returned radiation data.	Moon P1A
1960 Mar. 11	Pioneer 5 sun orbit.	Orbit between Earth and Venus. Proved theory of planetary magnetic fields w/ first map.	Sun P1A
1962 Dec. 14	Mariner 2 Venus fly-by.	First probe to reach another planet. Scans of atmosphere and surface. Later orbited sun.	Venus P1A
1962	Apollo Manned orbit.		Earth M1
1964 Feb. 2	Ranger 6 moon impact.	On target, schedule. Camera system fails.	Moon P2
1964 July 31	Ranger 7 moon impact.	On target. Sends back more than 4000 images during descent. Used to determine safe landing zones for Apollo.	
1965 July 15	Mariner 1 Mars fly-by.	Images of craters. Martian atmosphere thin, life unlikely. Reactivated 2 years later to support Mariner 5 to Venus	Mars P1A
1966 June 2	Surveyor 1 moon landing.	Survives Ocean of Storms landing, 14-m from target. 11,350 images used to design Apollo landers.	Moon P3A
1967 Nov. 17	Surveyor 6 moon lift-off.	First successful lift-off. Necessary for manned mission.	Moon P3B
1969 July 20	Manned Apollo moon mission.	Neil Armstrong and Buzz Aldrin.	Moon M3

Table B.1: DSN Mission Timeline 1958-1969

DATE	EVENT	NOTES	MISSION STAGE
1971 Nov. 14	Mariner 9 Mars orbit.	First probe to orbit another planet. Maps 85% of planet at 1-2 km resolution. Identifies Olympus Mons.	Mars P1B
1972 July	Pioneer 10 crosses asteroid belt.	Between Mars and Jupiter. Completes crossing February 1973.	Asteroid Belt P1A
1973 Dec. 4	Pioneer 10 Jupiter fly-by.	First probe to outer planets. 300 images.	Jupiter P1A
1974 Feb. 5	Mariner 10 gravity-assist.	First gravity-assist. Change trajectory at Venus to get to Mercury. Saves fuel, opens up outer system to exploration. Takes more than 4100 images of Venus.	
1974 Mar. 29	Mariner 10 Mercury fly-by.	Images. Battered, Moon-like surface. No atmosphere.	Mercury P1A
1976 July 20	Viking 1 Mars Landing.	First safe landing on Mars. Sends back first surface pictures. Studies soil samples. Regular weather reports.	Mars P3A
1979 Sept. 1	Pioneer 11 Saturn fly-by.		Saturn P1A

Table B.2: DSN Mission Timeline 1970-1979

DATE	EVENT	NOTES	MISSION STAGE
1985 Sept. 11	ICE flies past comet.	International Cometary Explorer confirms 'dirty snowballs' theory. Less dust than expected. Possible retrieval in 2014.	Comet P1A
1986 Jan. 24	Voyager 2 Uranus fly-by.	Photos. Discovered 10 new Uranian moons, two rings, and a boiling ocean of water about 800 km below cloud tops.	Uranus P1A
1989 Summer	Voyager 2 Neptune fly-by.		Neptune P1A

Table B.3: DSN Mission Timeline 1980-1989

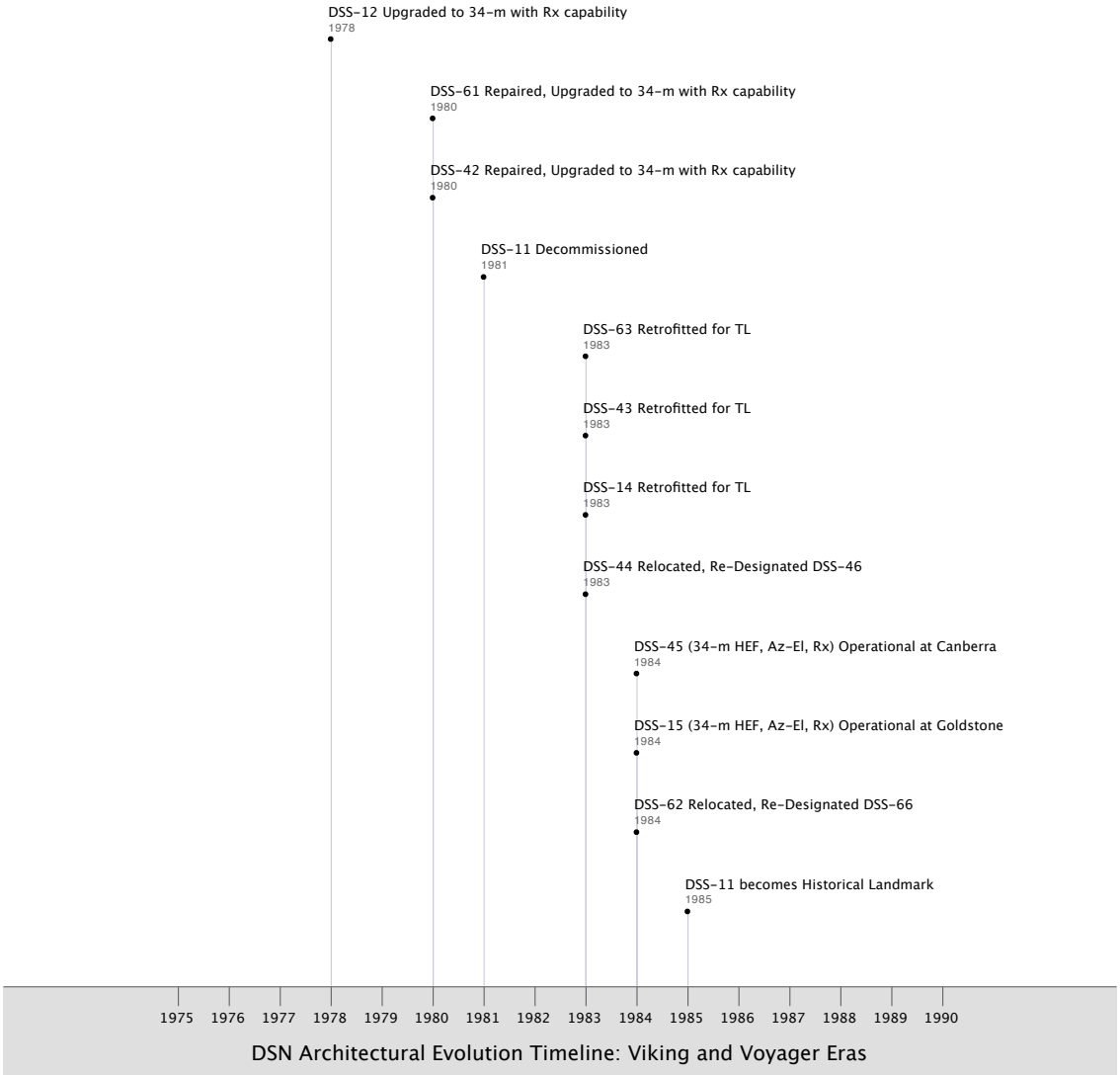
DATE	EVENT	NOTES	MISSION STAGE
1990	Magellan orbits Venus.	Reveals 85% of surface covered in lava. Deliberately destroyed in atmosphere in 1994.	Venus P1B
1990 Feb. 14	Voyager 1	Takes final images. Family portrait of solar system.	
1991 Oct. 29	Galileo asteroid fly-by.	First to fly past asteroid, Gaspara. 150 images.	Asteroid P1A
1993 Aug. 28	Galileo asteroid fly-by.	Flies past asteroid Ida, discovers first moon orbiting asteroid.	
1995	Galileo probe Jupiter impact.	Small probe descends into Jupiter's atmosphere. Survives 59 minutes. Spacecraft enters Jupiter orbit.	Jupiter P1B, P2
1997 July 5	Probe Surface Exploration.	Sojourner rover rolls off of Pathfinder onto surface of Mars.	Mars P3C

Table B.4: DSN Mission Timeline 1990-1999

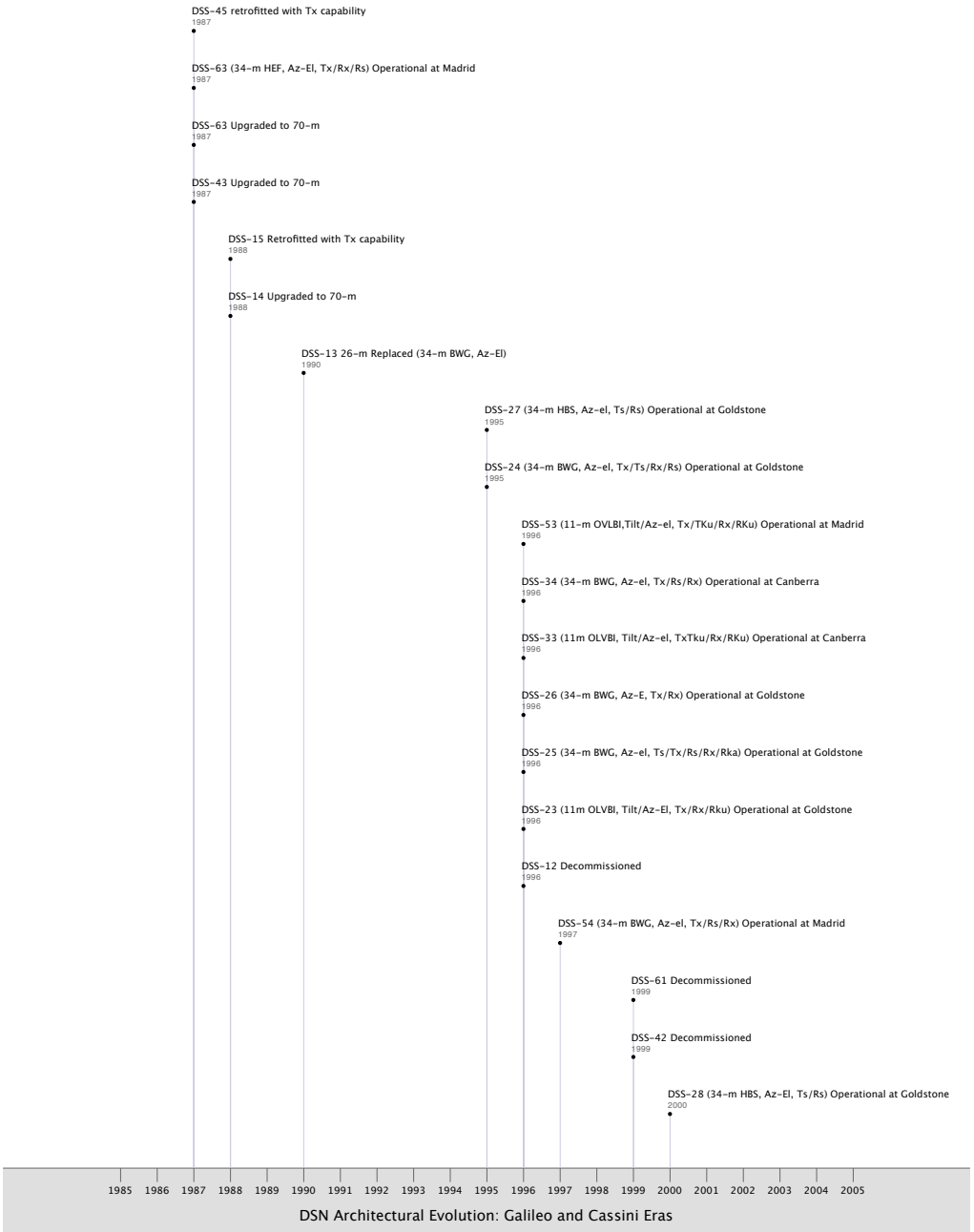
DATE	EVENT	NOTES	MISSION STAGE
2000 Feb. 14	NEAR orbits asteroid	NEAR spacecraft orbits asteroid, Eros.	Asteroid P1B
2001 Feb. 12	NEAR asteroid landing.	Successfully touches down on Eros. Not designed for landing. Attempt was opportunity at mission completion.	Asteroid P3A
2003 Nov. 5	Voyager 1 interstellar space.	Nears crossing into true interstellar space.	Interstellar P1A
2004 Jan. 2	Stardust comet fly-by.	Passes through gas and dust of comet Wild 2 and collects samples.	
2004 June 11	Cassini-Huygens Saturn fly-byt	Starts orbital tour of Saturn system.	Saturn P1B
2005 Jan. 14	Huygens probe Saturn landing.	Probe descends through Titan atmosphere, touching down.	Saturn P3A
2011	Messenger Mercury orbit.	Scheduled.	Mercury P1B
2015	New Horizons Pluto fly-by.	Scheduled.	Pluto P1A

Table B.5: DSN Mission Timeline 2000-2015

B.3 DSN Physical Architecture Evolution Timelines by Era



DSN Architectural Evolution Timeline: Viking and Voyager Eras


DSS-45 retrofitted with Tx capability
1987
DSS-63 (34-m HEF, Az-El, Tx/Rx/Rs) Operational at Madrid
1987
DSS-63 Upgraded to 70-m
1987
DSS-43 Upgraded to 70-m
1987
DSS-15 Retrofitted with Tx capability
1988
DSS-14 Upgraded to 70-m
1988
DSS-13 26-m Replaced (34-m BWG, Az-El)
1990
DSS-27 (34-m HBS, Az-el, Ts/Rs) Operational at Goldstone
1995
DSS-24 (34-m BWG, Az-el, Tx/Ts/Rx/Rs) Operational at Goldstone
1995
DSS-53 (11-m OVLBI,Tilt/Az-el, Tx/TKu/Rx/RKu) Operational at Madrid
1996
DSS-34 (34-m BWG, Az-el, Tx/Rs/Rx) Operational at Canberra
1996
DSS-33 (11m OLVBI, Tilt/Az-el, TxTku/Rx/RKu) Operational at Canberra
1996
DSS-26 (34-m BWG, Az-E, Tx/Rx) Operational at Goldstone
1996
DSS-25 (34-m BWG, Az-el, Ts/Tx/Rs/Rx/Rka) Operational at Goldstone
1996
DSS-23 (11m OLVBI, Tilt/Az-El, Tx/Rx/Rku) Operational at Goldstone
1996
DSS-12 Decommissioned
1996
DSS-54 (34-m BWG, Az-el, Tx/Rs/Rx) Operational at Madrid
1997
DSS-61 Decommissioned
1999
DSS-42 Decommissioned
1999
DSS-28 (34-m HBS, Az-El, Ts/Rs) Operational at Goldstone
2000
1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005
DSN Architectural Evolution: Galileo and Cassini Eras

B.4 DSN Snapshots of Interest

This section examines several well-bounded examples from the history of the DSN that provide crucial insights into the strategic evolution of systems. The first two snapshots concern the rehabilitation and upgrade of the 64-meter antennas, while the final snapshot details the evolution of the Mission Control and Computing Center.

In 1982, NASA commenced a project to revitalize the Deep Space Network. The project came in two parts: rehabilitating the aging 64-meter antennas, and upgrading them to achieve the performance improvement required to support missions to Neptune.[1]

B.4.1 Rehabilitating the 64-meter Antennas

The rehabilitation program was largely focused on the oldest 64-meter Deep Space Station (DSS) antenna, DSS-14 at the Goldstone complex. Over the years, the pedestal grout and concrete had deteriorated to critical levels, placing considerable stress on the hydrostatic bearing assembly. Similar issues were found in the radial bearing assembly.[1]

The remaining resources of the rehabilitation program went toward investigating a pedestal tilt at Canberra's DSS-43.[1]

The $3.8 million project repaired the DSS-14 hydrostatic bearing, replaced the DSS-14 radial bearing and investigated the cause of the DSS-43 pedestal tilt. The repairs at Goldstone were completed by June 1984, followed by the determination of the tilt in early 1985.[1]

Figure B.3 on page 320 provides a timeline of the critical periods and events during the rehabilitation of the 64-meter subnet.

The Design of the 64-meter Antenna Azimuth Rotation System

The DSN 64-meter antennas had an azimuth rotation system containing both hydrostatic and radial bearing assemblies. Both bearings are key features of the antenna design. Hydrostatic bearings support, guide and reduce the friction of motion of a shaft by virtue of high-pressure oil. The oil film raises and supports the shaft, enabling it to rotate smoothly. Radial bearings, on the other hand, support, guide and reduce the friction of motion of a shaft via rolling contact so the transmitted load

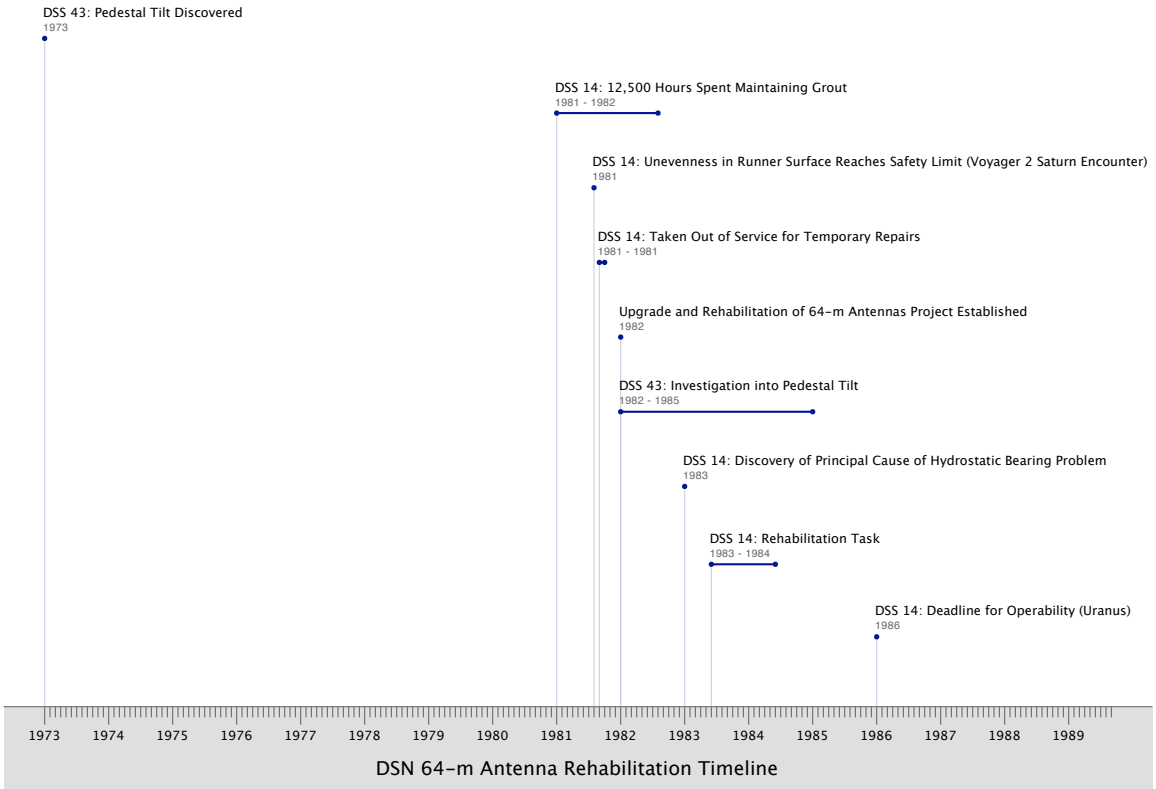


Figure B.3: DSN 64-m subnet rehabilitation timeline.

is radial to the axis of the shaft. The azimuth rotation system allows the antenna to rotate left and right to acquire and follow desired communication signals.[90,1]

The hydrostatic bearing assembly design is described in Uplink-Downlink as follows:

> *In this design, the 13.2-million-kilogram rotating structure was supported by a film of high pressure oil approximately 0.25 milimeter thick. The oil film formed the load-bearing medium between each of three "pads" that carried the entire rotating structure, and a large, horizontal circular steel "runner" bearing on which the shoes rotated in azimuth. The runner bearing had a finely machined upper surface to support the pads and was itself supported by a massive circular concrete pedestal. Between the upper surface of the concrete and the lower surface of the runner lay a thick layer of grout which was supposed to provide a stable, impervious, interface between the steel bearing and the concrete pedestal. The dimensions of the components of the hydrostatic bearing were impressive. Each pad was approximately 1 meter wide, 1.5 meter long, and 0.5 meter deep. The runner bearing was 24.4 meters in outside diameter, 1.12 meters wide,*

> *and 12.7 cm thick. The walls of the pedestal were more than one meter thick and were topped by a massive "haunch" 2.1 meters high and 1.8 meters thick. The haunch provided the foundation for the grout and the runner.*

The radial bearing assembly "consisted of a 30-foot-diameter steel ring surrounding a concrete collar at the top of the pedestal. A vertical wearstrip was attached to the runner to provide a track for three wheel assemblies equally spaced and attached to the rotating alidade structure. The wheel assemblies, bearing on the vertical wearstrip and runner, maintained the correct vertical axis of rotation for the entire antenna."[1]

Goldstone's DSS-14 Pedestal Deterioration

A consistent problem with the DSS-14 hydrostatic bearing assembly was deterioration of the concrete pedestal and the grout. As the concrete and the grout crumbled, the steel runners shifted, causing variations in the oil film height between the pads and runner bearings. Every time the oil film height dropped to a specified minimum value, the antenna was taken out of service for maintenance; the outage also meant a halt to mission operations as a whole. The problem became so bad that between January 1981 and August 1982, workers put in 12,500 hours of effort to maintain the grout, either by raising the critical section and inserting shims or by replacing the entire section of grout.[1]

By the time DSS-14 was set to make its final tracking pass of Voyager 2's encounter with Saturn in August 1981, the antenna hydrostatic bearing was consistently operating at or near the alarm level of 0.005 inches of film height. In an effort to stave off repairs until after the Saturn encounter, an engineer was posted on the antenna to monitor the oil film height. The bid was successful, and temporary repairs of the grout were made over the next weeks. It was now certain that a more permanent solution was required.[1]

Meanwhile, similar issues with the radial bearing assembly had elevatated the risk of failure to an unacceptably high level. In this case, the deterioration was occurring in the grout between the radial bearing runner and wearstrip. A failure of the radial bearing would require many months of repair.[1]

In 1983, years of study and analysis of the deterioration of the concrete and the grout paid off, with the discovery that the fundamental reason for the failure of the load-bearing grout was due to a chemical reaction in the concrete. The particular aggregate in the concrete mix used in the initial

construction of the antenna reacted with the alkali in the cement, forming silica gel. The silica "absorbed moisture, expanded, and caused microcracks in the concrete." The microcracks softened the concrete, leaving it unable to adequately support the hydrostatic bearing.[1]

DSS-14 Pedestal Repair

The DSS-14 antenna was removed from service in June 1983. The effort would take advantage of the planned downtime for repairing the pedestal concrete and refurbishing the hydrostatic assembly to simultaneously repair the radial bearing.

Uplink-Downlink describes the DSS-14 rehabilitation as follows:

> *The six-million-pound rotating structure was raised and placed on temporary supporting columns to allow the hydrostatic bearing and the runner to be removed for rework. All the concrete in the pedestal haunch was then removed one section at a time by a four-person crew with jackhammer and drills. Because of the enormous amount of reinforcing steel embedded in the concrete, it took the crew about five days to remove a 40-foot-long section. As each section was removed, the new concrete* — 450 cubic yards worth — *was poured and allowed to cure. When the new "haunch" was completed, the radial bearing was replaced and the hydrostatic bearing components installed and aligned. Final tests on the new type of concrete showed that the original specifications for stiffness had been met or exceeded.*[1]

The radial bearing replacement and refurbishment required the fabrication of a new runner and wearstrip. The installation process necessitated the adjustment and alignment of these parts with the center of the antenna foundation, requiring several weeks of effort. For maximum strength, the grout between the runner and wearstrip was replaced and allowed to cure for two weeks. The wheel assemblies were sent away for reconditioning. Their reinstallation produced a few hiccups due to difficulties finding the correct realignment. The workers successfully finished the radial bearing repair work in time to support "the first rotation of the antenna on the new pedestal."[1]

All told, the process of raising and temporarily supporting the antenna — in addition to removing and replacing the concrete, the hydrostatic bearing assembly, and the radial bearing assembly — took almost a year. Amazingly, the "task was completed on schedule, within budget, and without a lost-time accident."[1]

Canberra's DSS-43 Pedestal Tilt

The pedestal tilt at DSS-43 was discovered in 1973 and was believed to be due to nonuniform loading. The tilt was not considered an issue until 1982 when planning began for the expansion of the 64-meter antennas to 70-meter antennas and the impact of the additional weight was unknown.[1]

An investigation into the tilt revealed that the soil and bedrock underneath the south side of the pedestal was softer than the material supporting the north side. The many years of service of the antenna had completely compressed the soil, and any additional weight was therefore not expected to significantly increase the tilt.[1]

Analysis

There are several lessons to be learned from the rehabilitation programs for DSS-14 and DSS-43. The programs can be broken down into events, actions, support processes and decisions.

Analysis of DSS-14 Rehabilitation The critical deterioration of the DSS-14 bearing assemblies was an initial internal event. In particular, this event was a threat to the performance of the entire network. In response, the decision makers gave the go-ahead to maintain the grout for the hydrostatic bearing. This was merely a temporary measure, but until the parallel support process completed its study and analysis of the problem to find the fundamental cause, permanent repairs could not be done. A secondary internal threat event occured when the hydrostatic bearing deterioration reached the pre-defined safety limit. The cause of the problem had still not been found and the threat had occurred during a critical operational period of the network. The decision was made to simply monitor the condition until the mission was complete and then make temporary repairs afterwards. Finally, when the support process was successful and could provide an explanation for the deterioration and a plan to correct the problem, the decision makers gave the go ahead to rehabilitate and repair the antenna. During this entire time, the radial bearing assembly had also experienced deterioration, though not as severe. The decision to take the antenna offline to repair the hydrostatic bearing acted as an internal opportunity event. The decision was made to leverage the planned downtime to also repair the radial bearing deterioration.

Analysis of rehabilitation program for DSS-43 The DSS-43 pedestal tilt became an internal threat event when the decision was made to extend the diameter. Since the cause of the tilt was

unknown, engineers were uncertain what its structural response would be to the added weight. The support process, which had been charged with studying the problem, eventually came back with an identified cause and the projected impact. Based on the positive report, the decision makers decided to maintain the antenna "as is."

B.4.2 Upgrading the 64-meter Antennas

The project to upgrade the 64-meter antennas proceeded in combination with the rehabilitation program discussed in the previous section. The antennas required a performance boost to support the DSN's upcoming missions to Neptune. To achieve the desired performance, the program focused on three key areas: increasing the existing 64-m antenna diameter to 70-m, improving the stiffness of the structure by modifying the structural braces, and improving the subreflector focus capability by adding automatic Y-axis focusing.[1]

The decision to increase diameter of the existing antennas strongly motivated the rehabilitation efforts of DSS-14. In addition to the $3.8 million spent on the rehabilitation projects, the antenna upgrade task cost another $4.1 million.[1]

Figure B.4 provides a timeline of the critical periods and events during the upgrade project for the 64-meter subnet.

Motivations for Upgrading

The motivation for upgrading the 64-meter antennas came in the early 1980s, the Deep Space Network found itself underpowered for the upcoming Voyager 2 flyby's of Uranus in 1986 and Neptune in 1989. An Aperture Unit (Ap.U.) metric was developed to aid in evaluating strategies for boosting the necessary downlink performance. The Ap.U. metric was "equivalent to the effective aperture of a DSN 64-meter antenna with a system noise temperature of 25 kelvin at X-band." The Uranus encounter would require 5.25 Ap.U. while Neptune needed a whopping 10.92 Ap.U.[1]

A solution for the Uranus flyby was found by 1983. Of the 5.25 Ap.U., 4.45 Ap.U. would come from an array of DSN antennas. The remaining 0.8 Ap.U. would come from the 64-meter Parkes observatory antenna in Parkes, New South Wales, Australia.[1]

The Uranus configuration could be reused for the Neptune encounter, but there was still 5.67 Ap.U. unaccounted for.

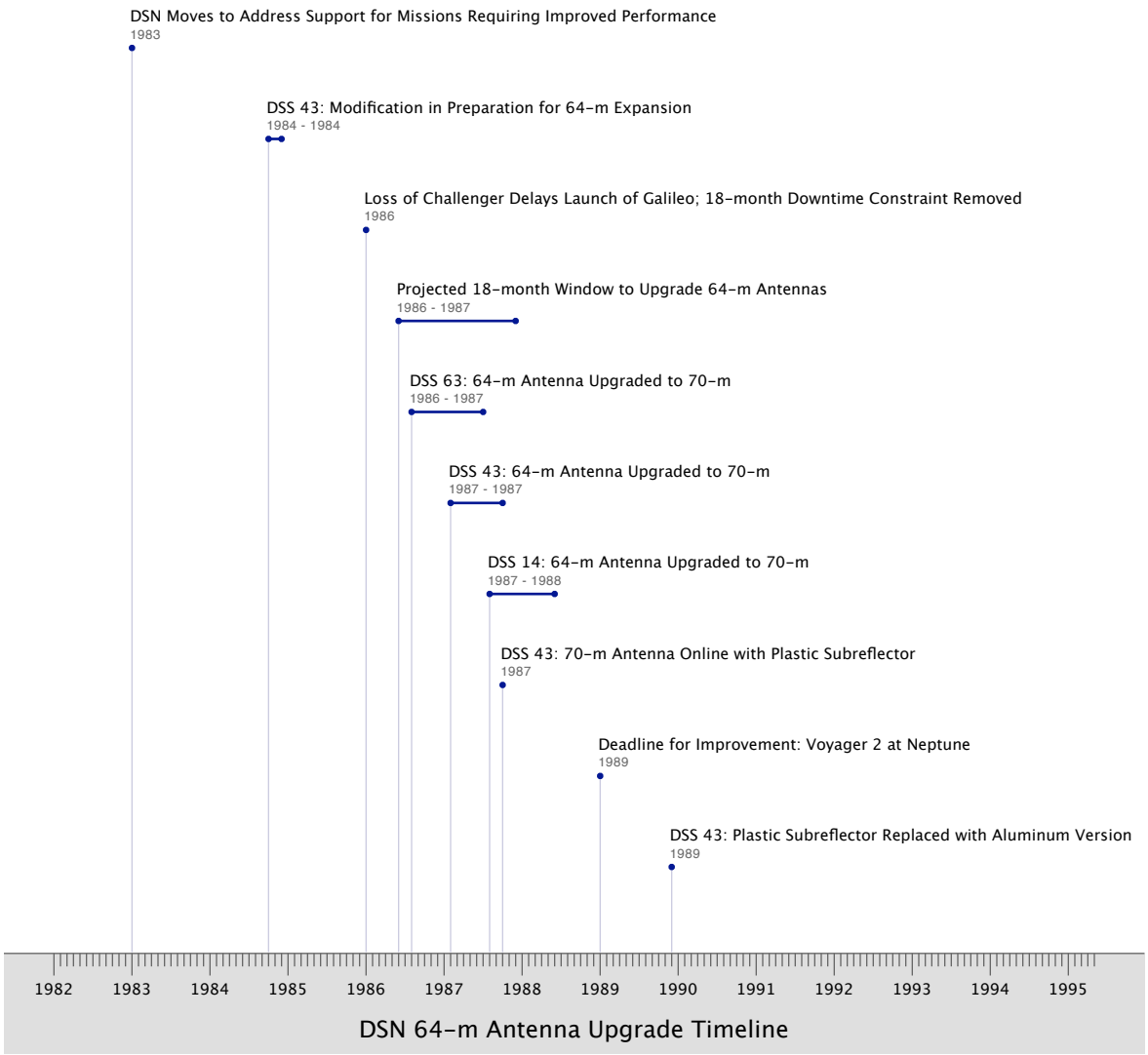


Figure B.4: DSN 64-m subnet upgrade project timeline.

Upgrading Options and Performance Analysis

This section briefly describes the upgrading options for the 64-meter antennas and the performance analysis undertaken before and after the upgrade. Figure B.5 on page 326 shows the upgrade options available to NASA along with the expected performance increases in X-band and S-band and the anticipated cost per antenna.

The run-up to the Voyager 2 flyby was not the first time NASA and JPL engineers had considered enhancing the 64-meter antenna performance. The idea had been analyzed for years, providing a multitude of documented options complete with performance, cost and viability information.

Options for Upgrading the 64-Meter Antenna

Upgrade Option	X-band Perf. Increase	S-band Perf. Increase	Cost per Antenna (FY 1982)	X-band Value in $M/dB[4]
Ultimate[1]	1.9 dB	1.4 dB	$6,000,000	3.2
Intermediate[2]	1.4 dB	1.4 dB	$3,300,000	2.4
Gravity[3] Compn.	0.3 dB	0.0	$330,000	1.1

1. Ultimate: Extend 64-m to 70-m; shaped precision reflectors; gravity compensation.
2. Intermediate: Extend 64-m to 70-m; shaped reflectors; gravity compensation.
3. Gravity compensation: Braces and Y-axis correction to 64-m or 70-m tipping structure.
4. For comparison, the arrayed addition of one new 34-m antenna to a standard 64-m antenna provides approximately 1.3 dB performance increase for a cost of $6.0M, as installed and equipped for X-band reception only. The equivalent value is $4.6M/dB. However, the added 34-m antenna would provide a capability for simultaneous multiuser support which is not given by the single, upgraded 64-m antenna option.[41]

Figure B.5: Options for upgrading the 64-m antennas. NASA chose the ultimate option. The table is taken from Uplink-Downlink.[1]

Robertson Stevens, an engineer who had been strongly involved with the establishment of the original 64-meter Goldstone antenna, was a major proponent for the enhancement project. He proposed, based on previous studies, to secure an additional 1.65 Ap.U. (over the DSN contribution to the Uranus configuration) by upgrading the existing DSN 64-meter subnet to 70-meter. An extra 0.35 Ap.U. could be achieved by augmenting the Madrid complex with a new high-efficiency antenna. Various non-NASA entities would provide the remaining 4.47 Ap.U. These outside facilities included: the 64-meter radio astronomy Parkes antenna in Australia, the Japanese Space Agency 64-meter antenna in Usuda, Japan, and the U.S. National Radio Astronomy Observatory very large array antennas in New Mexico.[1]

The driving motivation for upgrading the 64-meter antennas was the Voyager Neptune encounter. As Steven's proposal for meeting the needs of the Neptune flyby was not only feasible but cost-

effective, NASA decided to approve the plan.[1]

The upgrade project for the 64-meter antennas had many technical advantages in addition to the improvement in downlink antenna gain. These benefits included:[1]

- Increased gain applied to uplink as well, providing improved command capability under "adverse conditions."
- Increased sensitivity to aid in very long baseline interferometry VLBI, allowing a greater selection of radio sources.
- Improved carrier tracking of signals from smaller antennas during arrayed operational modes.
- Reduced cost per aperture unit of project relative to building a new 34-meter antenna. The extension to 70-meter was estimated to be only 60 percent of the cost of building a new 34-meter antenna.
- Similar operations and maintenance costs to existing 64-meter antennas, compared with about $200,000 per year of an additional 34-meter antenna.

The Implementation

This section describes the implementation of the 64-meter upgrade project.

Scheduling challenges meant that the transformation from the 64-meter subnet to 70-meter required a clever implementation strategy. It was expected that the task completion for each antenna would take approximately 12 months, or a total of three years if done consecutively. The mission schedule drove the available work window: the only opportunity appeared to be an 18-month period between the mid-1986 launch of Galileo and the spacecraft's probe release in late 1987. The project team needed to come up with a way to complete the modifications in half of the time.[1]

The team accomplished the reduced schedule by employing three time-saving strategies:[1]

- The new design would be performed primarily by JPL with assistance from TIW Systems. JPL would act as the prime contractor for the Madrid and Goldstone antennas; the Canberra antenna fabrication and erection would be contracted locally. The standard approach — with too large of a lead time for this situation — involved a contract with a single agent who would be responsible for everything.

- The 70-meter antennas would be brought online at a 64-meter antenna performance with final tuning to 70-meter performance as time permitted.
- The work schedule could be compressed by intelligently ordering the antenna modifications. Madrid would be first, followed by Australia and finally Goldstone. The drivers for this order included: ongoing mission support, site weather conditions, availability of equipment and subcontractors and the urgency of the Goldstone assignments.

Highlights of Madrid site implementation: Accumulated work delays extended the Madrid antenna downtime by eight weeks; all efforts to mititage the delays proved futile. Since all of the antenna conversions were planned to be serial, these delays presented a problem. In order to compress the newly projected time frame for task completion, it was decided that the work on the next site would be started on time despite the resulting concurrent downtime of the antennas. Based on this experience, the remaining site schedules were extended by two months.[1]

Further delays occurred during the alignment of the mid-December 1986 surface panels installation due to an unexpected 100-year winter storm.

Following the installation of the aluminum subreflector, it was discovered that the antenna performance was significantly deficient. The panels were readjusted, but the performance declined. A special "Tiger Team" was called in to investigate the problem. Finally, the team was successful in exceeding the performance of the 64-meter antennas by 2.1 dB after tightening the "kicker braces" and using holography to perform the panel alignment.

Highlights of Canberra site implementation: Different funding arrangements in Australia led to much of the work being contracted locally. Once all of the necessary material and equipment was ready and on-site, the antenna was brought off-line for the upgrading work.

The Canberra implementation experienced small delays caused by misalignment of the truss modules and issues with mounting a portion of the surface panels.

The problems with the Madrid panel alignment led the Canberra team to perform their final alignment settings using a new, time-consuming process known as holography. This decision was successful, but added three weeks to the completion date.

When the antenna modification was completed, a small decrease was measured in Canberra's antenna performance relative to the Madrid antenna. The difference was likely due to the use of a

temporary plastic subreflector. Since the antenna still met the 70-meter performance specifications, it was brought online and operational for two years before the permanent aluminum version was installed.

Highlights of Goldstone site implementation: The Goldstone antenna upgrade team benefited from the experiences at other sites. The alignment, installation and tuning processes went smoothly. Following the Canberra team, the tuning task was done first using rough setting, followed by the more precise holography.

Results

Overall, the 64-meter antenna upgrade project was quite successful. Using some creative techniques, the program team was able to meet the schedule, despite some unexpected hurdles. Measurements before and after the upgrade demonstrated that the expected performance was met and even exceeded at some sites.

The ambitious 18-month schedule was re-evaluated when the Galileo launch was delayed by three years due to the January 1986 loss of the Shuttle Challenger. Although the next mission constraint was the Neptune flyby in August 1989, it was decided that the project would stick to the original June 1988 subnet completion date. To achieve this milestone, it was necessary to overlap the downtime with the Madrid and Canberra antennas. The breakdown of the actual start and completion dates for each of the upgraded antennas is shown in Figure B.6 on page 330.[1]

Both of the Madrid and Canberra antennas exceeded the desired improvement in gain performance in X-band, as shown in Figure B.7 on page 330. Goldstone realized the performance requirements with an improvement of +1.84 dB in X-band. The differences in improvement between Goldstone and the other stations were eventually understood and corrected.

Analysis

Although the upgrade project followed the rehabilitation project, it was a major motivator for some of the rehabilitation work. Clearly, some changes occur in anticipation of future events or future changes, while others are made in response to events that have already happened. Even the upgrade

Summary of 64-m to 70-m Antenna Expansion Start and Finish Dates

Location	Start Date	Return to Operations with 70-m Performance
Madrid	1 Aug 1986	29 July 1987
Canberra	1 Feb 1987	21 Oct 1987
Goldstone	1 Aug 1987	1 June 1988

Figure B.6: 64-m antenna subnet expansion start dates and corresponding 70-m antenna subnet completion dates. Table taken from Uplink-Downlink.[1]

Summary of Measured Antenna Gain Before and After Upgrade

Gain (dBi)	DSS 14 X-band	DSS 43 X-band	DSS 63 X-band	All Ants. S-band
64-m	72.33	71.94	71.91	61.40
70-m	74.17	74.10	74.28	63.34
Improvement	+1.84	+2.16	+2.37	+1.94

Figure B.7: Measured antenna gains before and after upgrade project. Table taken from Uplink-Downlink.[1]

project was undertaken in anticipation of an increase in performance required to support a future mission to Neptune.

Extending the 64-meter antennas had been considered by NASA and JPL's support process engineers for years, largely due to the vision of some of their key people. The support process enabled the development of clever metrics, such as Ap.U., for evaluating the several proposals created for achieving their aims.

The decision makers were provided with several options by the support process, which considered modifications to legacy assets as well as the installation of new assets. At the time, it was considerably more cost-effective to modify the legacy infrastructure.

Some of the most impressive project management involved the scheduling issues involved in making three years worth of upgrades in an 18-month time frame. This feat was made possible by distributing responsibility for each of the complexes to local contractors, careful identification of issues and intelligent ordering of antenna modifications at each complex, and placing priority on making the upgrades necessary for returning the antenna to 64-meter performance before the final

tuning to fully convert to 70-meter.

Lessons learned from modifications to one complex were methodically applied to the later complexes to reduce problems and delays. Downtime was minimized by ensuring all materials and equipment were onsite before work began.

Unexpected deficiencies in performance triggered an internal threat event, and support processes and actions were undertaken to resolve the problem.

B.4.3 Evolution of the Mission Control and Computing Center

Here, the evolution of the Mission Control and Computing Center is described.

The ground data system of the DSN changed dramatically as the DSN itself changed. Early on, the data system supported multiple single missions by assigning flight projects to one of three International Business Machines Corporation (IBM) 7094 computers. Revolutions in computer architecture enabled the DSN to move toward multi-mission capability. Toward this end, the Mission Control and Computing Center MCCC was created with a core of IBM 360-75's for flight projects. This core system was shared in a multiprocessing mode. Realtime processing capability was phased in using minicomputers from 1973 to 1981. Non-realtime processing, such as application programs and data records, was completed on mainframe computers. The Flight Projects Support Office (FPSO) negotiated the usage of these resources with the flight projects.

The data processing on the MCCC/FPSO system was costly, so a facility utilizing state-of-the-art distributed processing technology was developed in the mid-1980's to reduce costs. The new facility was known as the Space Flight Operations Center (SFOC), and would perform all of the functions of the old MCCC. The MCCC was slowly transitioned over to the SFOC in the early 1990's, adapting new flight projects to the available multi-mission capabilities. Eventually, the SFOC became the core of the Advanced Multi-Mission Operations System (AMMOS), an "even more advanced data processing system."[1]

The evolution of the MCCC is shown in Figure B.8 on page 332.

This snapshot provides a good example of a build-new system being developed and phased in with concurrent operations. The development of distributed processing and workstation LAN also fundamentally acts as an external event (opportunity) that was leveraged to construct a build-new system.

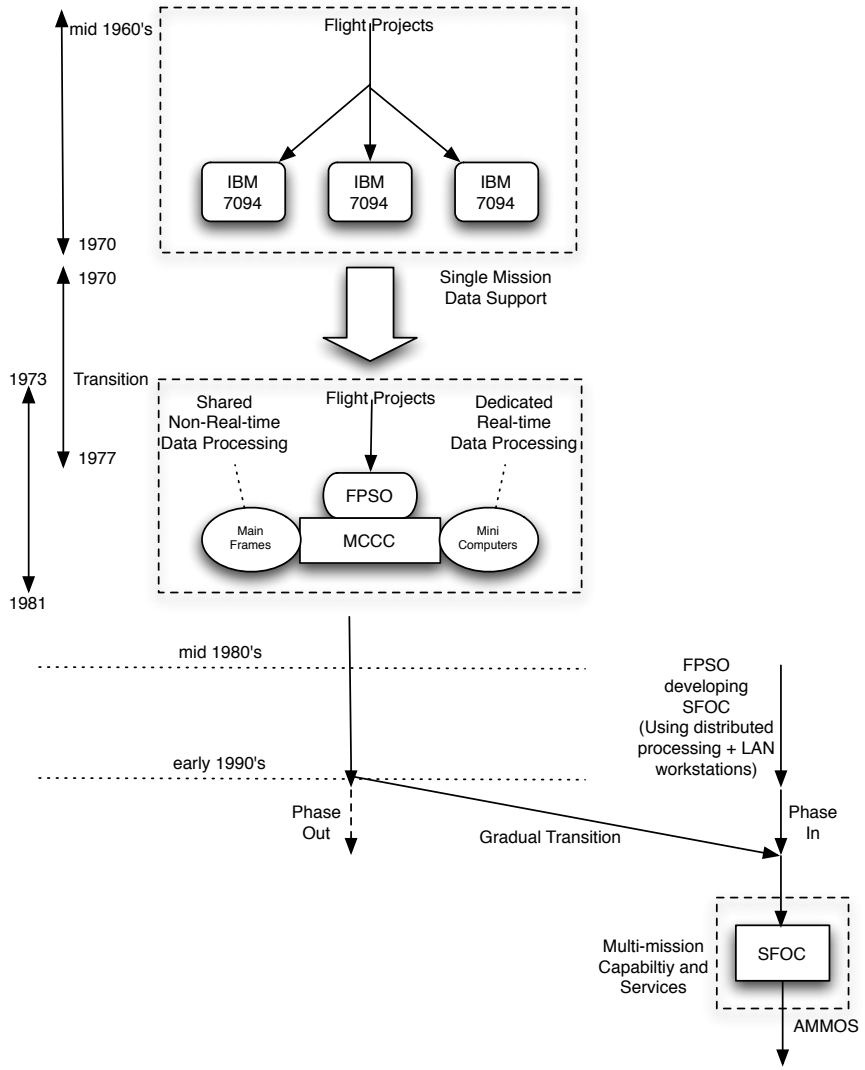


Figure B.8: Evolution of the Mission Control and Computing Center (MCCC).

B.4.4 Synopsis

This section has provided examples of: internal and external events; threats and opportunities; the role of support processes and decision makers; legacy systems, build-new systems and maintain 'as-is' systems; and change options and change mechanisms.

The Rehabilitation and Upgrade project of the 64-meter antennas provides ample examples of the role of the support processes and decision makers. The effect of legacy systems is apparent and there is even a case where the decision was to maintain "as-is." Examples of internal threats and

opportunities abound. Some relationships between change options and change mechanisms is also apparent.

The evolution of the MCCC is crucial as it provides an example of a build-new system and the resulting transition without interrupting operations. There is also a case of an external event driving the creation of a support process and decision.

Appendix C

Proofs of Position Entropy Theorems

Position entropy is a measure of the desirability (i.e., dynamic, multidimensional value) of an architecture and is defined as the sum of the entropy of the system of options out of the given architecture and the individual entropies of exercising those options, each weighted with the probability of exercising that particular option (refer to Figure C.1).

The α, β and γ factors are all assumed to be equal to one.

The position entropy can be written as:

$$H(i) = -\sum_{j\in k} p_{i\to j} \ln\left(\frac{1}{C(k)}\right) + \sum_{j\in k} p_{i\to j} \ln\left(\frac{E_i}{E_j}\right) + \sum_{j\in k} p_{i\to j} \ln\left(\frac{1}{E_j}\right) + \sum_{j\in k} p_{i\to j} H(j) \tag{C.1}$$

where the probability of exercising option j is:

$$p_{i\to j} = \frac{V_j}{\sum_{j\in k}(V_j)} \tag{C.2}$$

Such that the value of exercising option j is assumed to be:

$$V_j = \left(\frac{1}{E_j}\right) \tag{C.3}$$

and C(k) is the cardinality (number of elements) of the set of options, k.

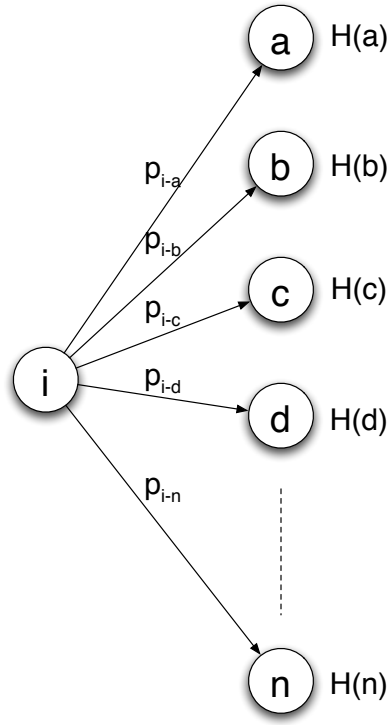


Figure C.1: Generic transition graph from an architecture i.

The position entropy is defined such that a larger entropy is more desirable. This definition represents the philosophy that the more desirable an architecture is, the more it is preferred. That is,

$$H(A) > H(B)$$
$$\Rightarrow H(A) \succ H(B)$$

Theorem 1 (*magis bene*): If architecture A has *more* transition options than architecture B, with all else equal, then architecture A has the preferred entropy to architecture B.

Theorem 2 (*melior bene*): If architecture A has *better* transition options than architecture B, with all else equal, then architecture A has the preferred entropy to architecture B.

Theorem 3 (*optimus bene*): Suppose there exist 2 paths leading to the same optimal (i.e., lowest energy, positive future entropy) architecture. If the first path achieves the optimal architecture in 1 transition, and the second achieves it in 2 transitions, such that the optimal architecture

is the only one with sub-unity energy, and with all else equal, the minimum transition path should contribute greater entropy to the current architecture.

Proof of Theorem 1: *Magis Bene*

If architecture A has *more* transition options (n) than architecture B (m), with all else equal, then architecture A has the preferred entropy to architecture B. Thus,

$$n > m$$

To prove Theorem 1, it is necessary to show that:

$$H(A) \succ H(B)$$
$$\Rightarrow H(A) > H(B)$$

It should be clear from looking at Figure C.2 that the generic position entropy equation

$$H(i) = -\sum_{j\in k} p_{i\to j} \ln\left(\frac{1}{C(k)}\right) + \sum_{j\in k} p_{i\to j} \ln\left(\frac{E_i}{E_j}\right) + \sum_{j\in k} p_{i\to j} \ln\left(\frac{1}{E_j}\right) + \sum_{j\in k} p_{i\to j} H(j)$$

becomes the following:

$$H(A) = -\sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n} p_{A\to j} H \tag{C.4}$$

$$H(B) = -\sum_{j=1}^{m} p_{B\to j} \ln\left(\frac{1}{m}\right) + \sum_{j=1}^{m} p_{B\to j} \ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{m} p_{B\to j} \ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{m} p_{B\to j} H \tag{C.5}$$

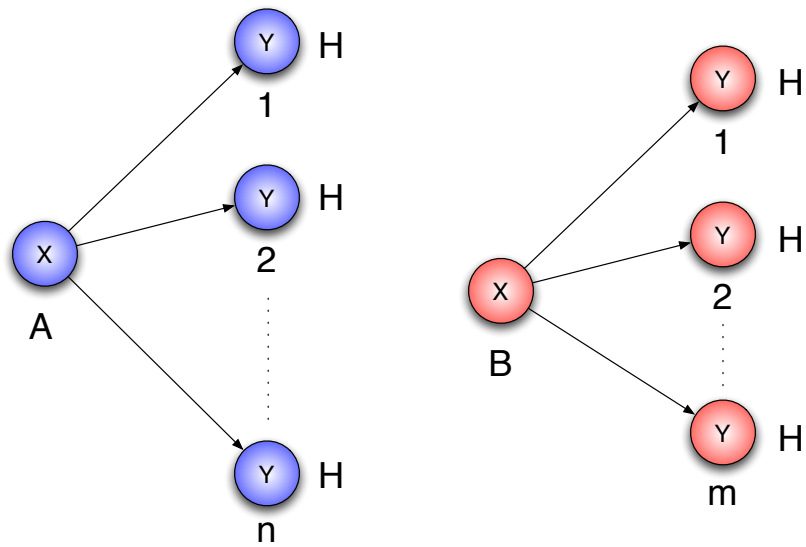


Figure C.2: Architecture A has more options (n) than Architecture B (m). Architecture A will have a greater (preferred) entropy value relative to Architecture B.

The probabilities of exercising the options for both architectures A and B are:

$$p_{A\to j} = \frac{V_j}{\sum_{j=1}^{n}(V_j)} \tag{C.6}$$

$$p_{B\to j} = \frac{V_j}{\sum_{j=1}^{m}(V_j)} \tag{C.7}$$

Plugging in, the probabilities of exercising the options become:

$$\begin{aligned} p_{A\to j} &= \frac{\left(\frac{1}{Y}\right)}{\sum_{j=1}^{n}\left(\frac{1}{Y}\right)} \\ &= \frac{\left(\frac{1}{Y}\right)}{n\left(\frac{1}{Y}\right)} \\ &= \frac{1}{n} \end{aligned} \tag{C.8}$$

$$\begin{aligned} p_{B\to j} &= \frac{\left(\frac{1}{Y}\right)}{\sum_{j=1}^{m}\left(\frac{1}{Y}\right)} \\ &= \frac{\left(\frac{1}{Y}\right)}{m\left(\frac{1}{Y}\right)} \\ &= \frac{1}{m} \end{aligned} \tag{C.9}$$

Substituting these expressions into the entropy equations gives:

$$H(A) = -\sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)H \tag{C.10}$$

$$H(B) = -\sum_{j=1}^{m}\left(\frac{1}{m}\right)\ln\left(\frac{1}{m}\right) + \sum_{j=1}^{m}\left(\frac{1}{m}\right)\ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{m}\left(\frac{1}{m}\right)\ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{m}\left(\frac{1}{m}\right)H \tag{C.11}$$

Since:

$$\begin{aligned}\sum_{j=1}^{n}\left(\frac{1}{n}\right)\alpha &= \underbrace{\left(\frac{1}{n}\right)\alpha}_{1} + \underbrace{\left(\frac{1}{n}\right)\alpha}_{2} + \ldots + \underbrace{\left(\frac{1}{n}\right)\alpha}_{n} \\ &= n\left(\frac{1}{n}\right)\alpha \\ &= \alpha \end{aligned} \tag{C.12}$$

Simplifying H(A) and H(B) using the above identity gives:

$$H(A) = -\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H \tag{C.13}$$

$$H(B) = -\ln\left(\frac{1}{m}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H \tag{C.14}$$

To prove Theorem 1 it is necessary to show that the following statement is true:

$$H(A) > H(B)$$

By substituting the expressions for H(A) and H(B) into the above statement and showing that it always holds true given the conditions of the theorem, the theorem will be proven. Thus,

$$-\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H > -\ln\left(\frac{1}{m}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H \tag{C.15}$$

Eliminating the identical components gives the following statement:

$$-\ln\left(\frac{1}{n}\right) > -\ln\left(\frac{1}{m}\right) \tag{C.16}$$

To show that the above statement holds true, consider the statement given in the theorem statement:

$$m < n$$

The following important relationships can be derived:

$$\left(\frac{1}{n}\right) < \left(\frac{1}{m}\right) \tag{C.17}$$

$$\Rightarrow \ln\left(\frac{1}{n}\right) < \ln\left(\frac{1}{m}\right) \tag{C.18}$$

$$\Rightarrow -\ln\left(\frac{1}{n}\right) > -\ln\left(\frac{1}{m}\right) \tag{C.19}$$

The last relationship is identical to the statement needing to be proven.

Thus:

$$H(A) > H(B)$$

$$\Rightarrow H(A) \succ H(B)$$

Q.E.D.

Proof of Theorem 2: *Melior Bene*

If architecture A has *better* transition options than architecture B, with all else equal, then architecture A has the preferred entropy to architecture B.

There are three different ways to describe *better* transition options. Thus, there are three cases that need to be proven.

Case 1: Suppose A and B have the same current energy X and future entropies H. Then in order for A to have better transition options, the architectures immediately accessible from A need more preferable energies (Y) than the energies that the architectures immediately accessible from B have (Z).

Case 2: Suppose A and B have the same current energy X AND their immediately accessible architectures have the same energies (Y). Then in order for A to have better transition options, the future entropies of the architectures immediately accessible from A must be more preferable to the future entropies of the architectures immediately accessible from B.

Case 3: Suppose A and B share the same ratio of current energies (X and S, respectively) to energies of immediately accessible options (Y and T, respectively), and future entropies H. Then in order for A to have better transition options, the current energy of A (X) must be preferable to the current energy of B (S).

For all three cases, it is necessary to show that:

$$H(A) \succ H(B)$$

Case 1: *Different Transition Ratios, Same Current Architecture Energies*

Suppose A and B have the same current energy X and future entropies H (see Figure C.3). Then in order for A to have better transition options, the architectures immediately accessible from A need more preferable energies (Y) than the energies that the architectures immediately accessible from B have (Z). Thus,

$$Z > Y$$
$$\Rightarrow Y \succ Z$$

To prove Theorem 2, it is necessary to show that:

$$H(A) \succ H(B)$$
$$\Rightarrow H(A) > H(B)$$

It should be clear from looking at the figure below that the generic position entropy equation

$$H(i) = -\sum_{j\in k} p_{i\to j} \ln\left(\frac{1}{C(k)}\right) + \sum_{j\in k} p_{i\to j} \ln\left(\frac{E_i}{E_j}\right) + \sum_{j\in k} p_{i\to j} \ln\left(\frac{1}{E_j}\right) + \sum_{j\in k} p_{i\to j} H(j)$$

becomes the following:

$$H(A) = -\sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n} p_{A\to j} H \tag{C.20}$$

$$H(B) = -\sum_{j=1}^{n} p_{B\to j} \ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n} p_{B\to j} \ln\left(\frac{X}{Z}\right) + \sum_{j=1}^{n} p_{B\to j} \ln\left(\frac{1}{Z}\right) + \sum_{j=1}^{n} p_{B\to j} H \tag{C.21}$$

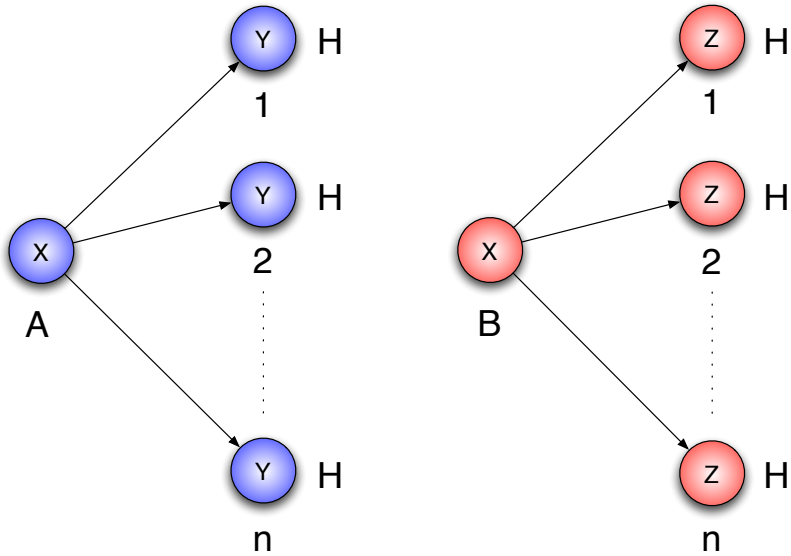


Figure C.3: Architecture A has better options (Y) than Architecture B (Z). Architecture A will have a greater (preferred) entropy value relative to Architecture B.

The probabilities of exercising the options for both architectures A and B are:

$$p_{A\to j} = \frac{V_j}{\sum_{j=1}^{n}(V_j)} \tag{C.22}$$

$$p_{B\to j} = \frac{V_j}{\sum_{j=1}^{n}(V_j)} \tag{C.23}$$

Plugging in, the probabilities of exercising the options become:

$$\begin{aligned} p_{A\to j} &= \frac{\left(\frac{1}{Y}\right)}{\sum_{j=1}^{n}\left(\frac{1}{Y}\right)} \\ &= \frac{\left(\frac{1}{Y}\right)}{n\left(\frac{1}{Y}\right)} \\ &= \frac{1}{n} \end{aligned} \tag{C.24}$$

$$\begin{aligned} p_{B\to j} &= \frac{\left(\frac{1}{Z}\right)}{\sum_{j=1}^{n}\left(\frac{1}{Z}\right)} \\ &= \frac{\left(\frac{1}{Z}\right)}{n\left(\frac{1}{Z}\right)} \\ &= \frac{1}{n} \end{aligned} \tag{C.25}$$

Substituting these expressions into the entropy equations gives:

$$H(A) = -\sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)H \tag{C.26}$$

$$H(B) = -\sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{X}{Z}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{Z}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)H \tag{C.27}$$

Since:

$$\begin{aligned}\sum_{j=1}^{n}\left(\frac{1}{n}\right)\alpha &= \underbrace{\left(\frac{1}{n}\right)\alpha}_{1} + \underbrace{\left(\frac{1}{n}\right)\alpha}_{2} + \ldots + \underbrace{\left(\frac{1}{n}\right)\alpha}_{n} \\ &= n\left(\frac{1}{n}\right)\alpha \\ &= \alpha \end{aligned} \tag{C.28}$$

Simplifying H(A) and H(B) using the above identity gives:

$$H(A) = -\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H \tag{C.29}$$

$$H(B) = -\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Z}\right) + \ln\left(\frac{1}{Z}\right) + H \tag{C.30}$$

To prove Theorem 2 it is necessary to show that the following statement is true:

$$H(A) > H(B)$$

By substituting the expressions for H(A) and H(B) into the above statement and showing that it always holds true given the conditions of the theorem, the theorem will be proven. Thus,

$$-\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H > -\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Z}\right) + \ln\left(\frac{1}{Z}\right) + H \tag{C.31}$$

Eliminating the identical components gives:

$$\Rightarrow \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) > \ln\left(\frac{X}{Z}\right) + \ln\left(\frac{1}{Z}\right) \tag{C.32}$$

Expanding the logarithms and combining terms gives:

$$\ln(X) - \ln(Y) - \ln(Y) > \ln(X) - \ln(Z) - \ln(Z) \tag{C.33}$$

$$-2\ln(Y) > -2\ln(Z) \tag{C.34}$$

$$\ln(Y) < \ln(Z) \tag{C.35}$$

Is this expression true?

The current case of the theorem provides that

$$Y < Z \tag{C.36}$$

Applying logarithms to this expression gives

$$\ln(Y) < \ln(Z) \tag{C.37}$$

which verifies the expression to be proven.

Therefore:

$$H(A) > H(B)$$

$$\Rightarrow H(A) \succ H(B)$$

Q.E.D.

Case 2: *Different Future Entropies, Same Architecture Energies*

Suppose A and B have the same current energy X AND their immediately accessible architectures have the same energies (Y) (see Figure C.4). Then in order for A to have better transition options, the future entropies of the architectures immediately accessible from A must be more preferable to the future entropies of the architectures immediately accessible from B. Thus,

$$H'(A) > H'(B)$$

To prove Theorem 2, it is necessary to show that:

$$H(A) \succ H(B)$$
$$\Rightarrow H(A) > H(B)$$

It should be clear from looking at the figure below that the generic position entropy equation

$$H(i) = -\sum_{j\in k} p_{i\to j} \ln\left(\frac{1}{C(k)}\right) + \sum_{j\in k} p_{i\to j} \ln\left(\frac{E_i}{E_j}\right) + \sum_{j\in k} p_{i\to j} \ln\left(\frac{1}{E_j}\right) + \sum_{j\in k} p_{i\to j} H(j)$$

becomes the following:

$$H(A) = -\sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n} p_{A\to j} H'(A) \qquad \text{(C.38)}$$

$$H(B) = -\sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n} p_{B\to j} \ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n} p_{B\to j} \ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n} p_{B\to j} H'(B) \qquad \text{(C.39)}$$

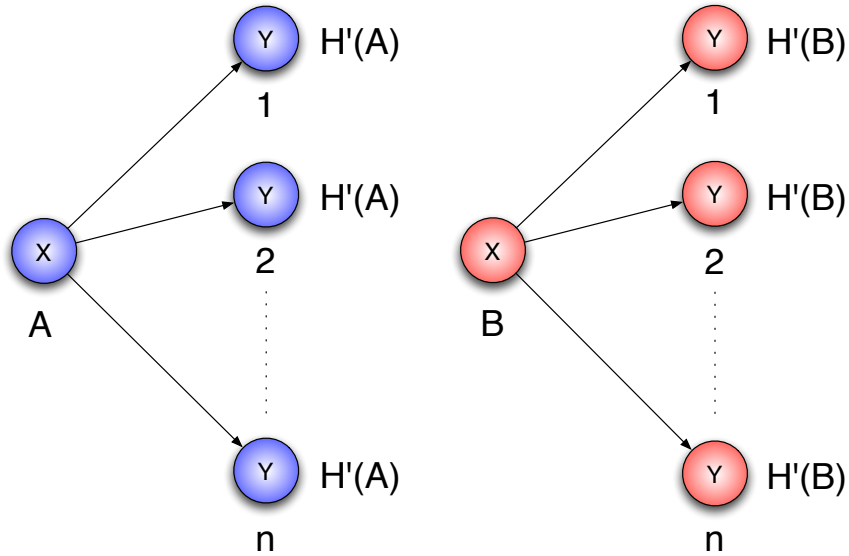


Figure C.4: Architecture A has better future options (H'(A)) than Architecture B (H'(B)). Architecture A will have a greater (preferred) entropy value relative to Architecture B.

The probabilities of exercising the options for both architectures A and B become:

$$p_{A\to j} = \frac{V_j}{\sum_{j=1}^{n}(V_j)} \tag{C.40}$$

$$p_{B\to j} = \frac{V_j}{\sum_{j=1}^{n}(V_j)} \tag{C.41}$$

Plugging in, the probabilities of exercising the options become:

$$\begin{aligned} p_{A\to j} &= \frac{\left(\frac{1}{Y}\right)}{\sum_{j=1}^{n}\left(\frac{1}{Y}\right)} \\ &= \frac{\left(\frac{1}{Y}\right)}{n\left(\frac{1}{Y}\right)} \\ &= \frac{1}{n} \end{aligned} \tag{C.42}$$

$$\begin{aligned} p_{A\to j} &= \frac{\left(\frac{1}{Y}\right)}{\sum_{j=1}^{n}\left(\frac{1}{Y}\right)} \\ &= \frac{\left(\frac{1}{Y}\right)}{n\left(\frac{1}{Y}\right)} \\ &= \frac{1}{n} \end{aligned} \tag{C.43}$$

Substituting these expressions into the entropy equations gives:

$$H(A) = -\sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)H'(A) \tag{C.44}$$

$$H(B) = -\sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)H'(B) \tag{C.45}$$

Since:

$$\begin{aligned}\sum_{j=1}^{n}\left(\frac{1}{n}\right)\alpha &= \underbrace{\left(\frac{1}{n}\right)\alpha}_{1} + \underbrace{\left(\frac{1}{n}\right)\alpha}_{2} + \ldots + \underbrace{\left(\frac{1}{n}\right)\alpha}_{n} \\ &= n\left(\frac{1}{n}\right)\alpha \\ &= \alpha \end{aligned} \tag{C.46}$$

Simplifying H(A) and H(B) using the above identity gives:

$$H(A) = -\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H'(A) \tag{C.47}$$

$$H(B) = -\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H'(B) \tag{C.48}$$

To prove Theorem 2 it is necessary to show that the following statement is true:

$$H(A) > H(B)$$

By substituting the expressions for H(A) and H(B) into the above statement and showing that it always holds true given the conditions of the theorem, the theorem will be proven. Thus,

$$-\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H'(A) > -\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H'(B) \tag{C.49}$$

Eliminating the identical components gives:

$$\Rightarrow H'(A) > H'(B) \tag{C.50}$$

Which is identical to the condition provided in the theorem case statement.

Therefore:

$$H(A) > H(B)$$

$$\Rightarrow H(A) \succ H(B)$$

Q.E.D.

Case 3: *Different Current Architecture Energies, Same Transition Ratios*

Suppose A and B share the same ratio of current energies (X and S, respectively) to energies of immediately accessible options (Y and T, respectively), and future entropies H (see Figure C.5). Then in order for A to have better transition options, the current energy of A (X) must be preferable to the current energy of B (S). Thus,

$$\left(\frac{X}{Y}\right) = \left(\frac{S}{T}\right)$$

It is given that:

$$X \succ S$$

$$\Rightarrow X < S$$

$$\Rightarrow Y < T$$

To prove Theorem 2, it is necessary to show that:

$$H(A) \succ H(B)$$
$$\Rightarrow H(A) > H(B)$$

It should be clear from looking at the figure below that the generic position entropy equation

$$H(i) = -\sum_{j \in k} p_{i \to j} \ln\left(\frac{1}{C(k)}\right) + \sum_{j \in k} p_{i \to j} \ln\left(\frac{E_i}{E_j}\right) + \sum_{j \in k} p_{i \to j} \ln\left(\frac{1}{E_j}\right) + \sum_{j \in k} p_{i \to j} H(j)$$

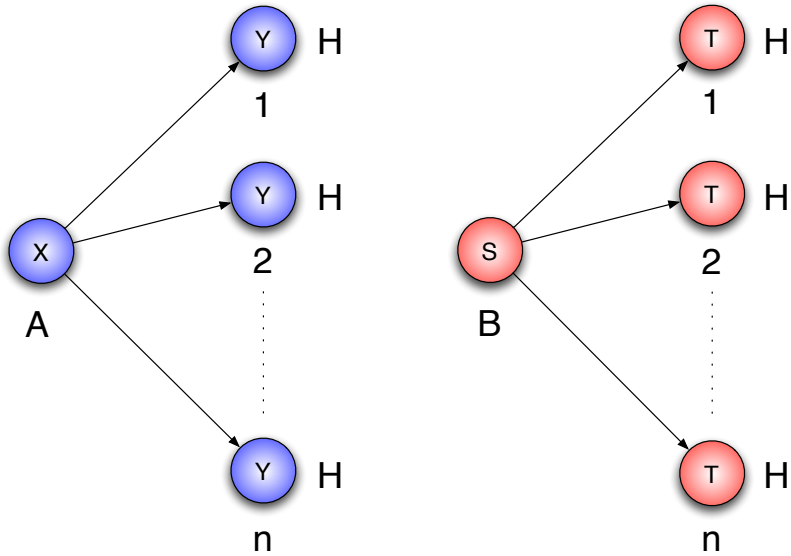


Figure C.5: Architecture A has better current energy (X) than Architecture B (S). Both architectures have the same transition ratio. Architecture A will have a greater (preferred) entropy value relative to Architecture B.

becomes the following:

$$H(A) = -\sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n} p_{A\to j} H \tag{C.51}$$

$$H(B) = -\sum_{j=1}^{n} p_{A\to j} \ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n} p_{B\to j} \ln\left(\frac{S}{T}\right) + \sum_{j=1}^{n} p_{B\to j} \ln\left(\frac{1}{T}\right) + \sum_{j=1}^{n} p_{B\to j} H \tag{C.52}$$

The probabilities of exercising the options for both architectures A and B are:

$$p_{A\to j} = \frac{V_j}{\sum_{j=1}^{n} (V_j)} \tag{C.53}$$

$$p_{B\to j} = \frac{V_j}{\sum_{j=1}^{n} (V_j)} \tag{C.54}$$

Plugging in, the probabilities of exercising the options become:

$$p_{A\to j} = \frac{\left(\frac{1}{Y}\right)}{\sum_{j=1}^{n} \left(\frac{1}{Y}\right)} = \frac{\left(\frac{1}{Y}\right)}{n\left(\frac{1}{Y}\right)} = \frac{1}{n} \tag{C.55}$$

$$p_{B\to j} = \frac{\left(\frac{1}{T}\right)}{\sum_{j=1}^{n}\left(\frac{1}{T}\right)} = \frac{\left(\frac{1}{T}\right)}{n\left(\frac{1}{T}\right)} = \frac{1}{n} \tag{C.56}$$

Substituting these expressions into the entropy equations gives:

$$H(A) = -\sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{X}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{Y}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)H \tag{C.57}$$

$$H(B) = -\sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{n}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{S}{T}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)\ln\left(\frac{1}{T}\right) + \sum_{j=1}^{n}\left(\frac{1}{n}\right)H \tag{C.58}$$

Since:

$$\begin{aligned}\sum_{j=1}^{n}\left(\frac{1}{n}\right)\alpha &= \underbrace{\left(\frac{1}{n}\right)\alpha}_{1} + \underbrace{\left(\frac{1}{n}\right)\alpha}_{2} + \ldots + \underbrace{\left(\frac{1}{n}\right)\alpha}_{n} \\ &= n\left(\frac{1}{n}\right)\alpha \\ &= \alpha\end{aligned} \tag{C.59}$$

Simplifying H(A) and H(B) using the above identity gives:

$$H(A) = -\ln\left(\frac{1}{n}\right) + \ln\left(\frac{X}{Y}\right) + \ln\left(\frac{1}{Y}\right) + H \tag{C.60}$$

$$H(B) = -\ln\left(\frac{1}{n}\right) + \ln\left(\frac{S}{T}\right) + \ln\left(\frac{1}{T}\right) + H \tag{C.61}$$

To prove Theorem 2 it is necessary to show that the following statement is true:

$$H(A) > H(B)$$

By substituting the expressions for H(A) and H(B) into the above statement and showing that it

always holds true given the conditions of the theorem, the theorem will be proven. Thus,

$$-\ln\left(\frac{1}{n}\right)+\ln\left(\frac{X}{Y}\right)+\ln\left(\frac{1}{Y}\right)+H>-\ln\left(\frac{1}{n}\right)+\ln\left(\frac{S}{T}\right)+\ln\left(\frac{1}{T}\right)+H \tag{C.62}$$

$$\Rightarrow -\ln\left(\frac{1}{n}\right)+\ln\left(\frac{X}{Y}\right)+\ln\left(\frac{1}{Y}\right)+H>-\ln\left(\frac{1}{n}\right)+\ln\left(\frac{X}{Y}\right)+\ln\left(\frac{1}{T}\right)+H \tag{C.63}$$

Eliminating the identical components gives:

$$\ln\left(\frac{1}{Y}\right)>\ln\left(\frac{1}{T}\right) \tag{C.64}$$

Returning to the theorem condition:

$$T>Y \tag{C.65}$$

$$\left(\frac{1}{Y}\right)>\left(\frac{1}{T}\right) \tag{C.66}$$

$$\ln\left(\frac{1}{Y}\right)>\ln\left(\frac{1}{T}\right) \tag{C.67}$$

Which verifies the expression to be proven.

Therefore:

$$H(A)>H(B)$$

$$\Rightarrow H(A)\succ H(B)$$

Q.E.D.

Proof of Theorem 3: *Optimus Bene*

Suppose there exist 2 paths leading to the same optimal (i.e., lowest energy, positive future entropy) architecture. If the first path achieves the optimal architecture in 1 transition, and the second achieves it in 2 transitions, such that the optimal architecture is the only one with sub-unity energy, and with all else equal, the minimum transition path should contribute greater entropy to the current architecture.

It should be clear from looking at Figure C.6 below that the entropy equations become:

$$\begin{aligned} H_c(X \to Z \to Z) = & -p_{X\to Z} \ln\left(\frac{1}{2}\right) + p_{X\to Z} \ln\left(\frac{X}{Z}\right) + p_{X\to Z} \ln\left(\frac{1}{Z}\right) \\ & + p_{X\to Z}\left[-p \ln p' + p \ln\left(\frac{Z}{Z}\right) + p \ln\left(\frac{1}{Z}\right) + pH\right] \end{aligned} \tag{C.68}$$

$$\begin{aligned} H_c(X \to Y \to Z) = & -p_{X\to Y} \ln\left(\frac{1}{2}\right) + p_{X\to Y} \ln\left(\frac{X}{Y}\right) + p_{X\to Y} \ln\left(\frac{1}{Y}\right) \\ & + p_{X\to Y}\left[-p \ln p' + p \ln\left(\frac{Y}{Z}\right) + p \ln\left(\frac{1}{Z}\right) + pH\right] \end{aligned} \tag{C.69}$$

It should also be clear that the probabilities of exercising the two options become:

$$p_{X\to Z} = \frac{\left(\frac{1}{Z}\right)}{\left(\frac{1}{Y}\right) + \left(\frac{1}{Z}\right)} = \frac{Y}{Y+Z} \tag{C.70}$$

$$p_{X\to Y} = \frac{\left(\frac{1}{Y}\right)}{\left(\frac{1}{Y}\right) + \left(\frac{1}{Z}\right)} = \frac{Z}{Y+Z} \tag{C.71}$$

Note: p' is the cardinality of the set of options in the second transition. Given the condition that all else be the same, p' is identical for both paths.

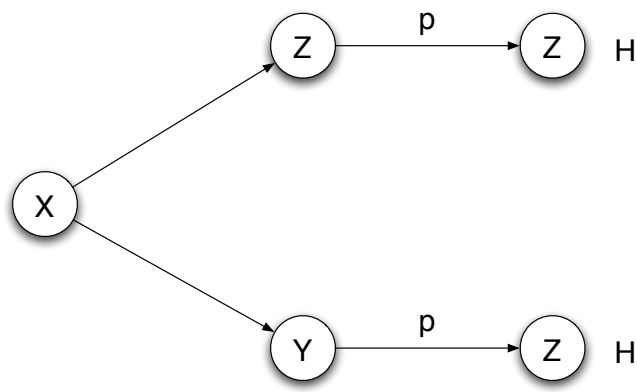


Figure C.6: Theorem 3: The path of X-Z-Z should contribute greater entropy than the path of X-Y-Z. The theorem assumes that all other contributions to the entropy are the same, thus the position entropy of Z on path 1 is greater than the position entropy of Y on path 2.

Substituting these expressions into the entropy equations gives:

$$\begin{aligned} H_c(X \to Z \to Z) = &-\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{1}{2}\right) + \left(\frac{Y}{Y+Z}\right)\ln\left(\frac{X}{Z}\right) + \left(\frac{Y}{Y+Z}\right)\ln\left(\frac{1}{Z}\right) \\ &-\left(\frac{Y}{Y+Z}\right)p\ln p' + \left(\frac{Y}{Y+Z}\right)p\ln\left(\frac{Z}{Z}\right) + \left(\frac{Y}{Y+Z}\right)p\ln\left(\frac{1}{Z}\right) + \left(\frac{Y}{Y+Z}\right)pH \end{aligned}$$

$$\begin{aligned} H_c(X \to Y \to Z) = &-\left(\frac{Z}{Y+Z}\right)\ln\left(\frac{1}{2}\right) + \left(\frac{Z}{Y+Z}\right)\ln\left(\frac{X}{Y}\right) + \left(\frac{Z}{Y+Z}\right)\ln\left(\frac{1}{Y}\right) \\ &-\left(\frac{Z}{Y+Z}\right)p\ln p' + \left(\frac{Z}{Y+Z}\right)p\ln\left(\frac{Y}{Z}\right) + \left(\frac{Z}{Y+Z}\right)p\ln\left(\frac{1}{Z}\right) + \left(\frac{Z}{Y+Z}\right)pH \end{aligned}$$

In order to prove Theorem 3, it is necessary to show that:

$$H_c(X \to Z \to Z) > H_c(X \to Y \to Z)$$

It will be shown that each pair of entropy contributions favors the first path over the second.

Thus, the following statements will be proven:

Statement 1: *First Transition Configuration Entropy Contribution*

$$-\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{1}{2}\right) > -\left(\frac{Z}{Y+Z}\right)\ln\left(\frac{1}{2}\right) \tag{C.72}$$

Statement 2: *Transition Entropy Contribution*

$$\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{X}{Z}\right)+\left(\frac{Y}{Y+Z}\right)p\ln\left(\frac{Z}{Z}\right) > \left(\frac{Z}{Y+Z}\right)\ln\left(\frac{X}{Y}\right)+\left(\frac{Z}{Y+Z}\right)p\ln\left(\frac{Y}{Z}\right) \tag{C.73}$$

Statement 3: *First Transition New State Entropy Contribution*

$$\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{1}{Z}\right) > \left(\frac{Z}{Y+Z}\right)\ln\left(\frac{1}{Y}\right) \tag{C.74}$$

Statement 4: *Second Transition Configuration Entropy Contribution*

$$-\left(\frac{Y}{Y+Z}\right)p\ln p' > -\left(\frac{Z}{Y+Z}\right)p\ln p' \tag{C.75}$$

Statement 5: *Second Transition New State Entropy Contribution*

$$\left(\frac{Y}{Y+Z}\right)p\ln\left(\frac{1}{Z}\right) > \left(\frac{Z}{Y+Z}\right)p\ln\left(\frac{1}{Z}\right) \tag{C.76}$$

Statement 6: *Future Entropy Contribution*

$$\left(\frac{Y}{Y+Z}\right)pH > \left(\frac{Z}{Y+Z}\right)pH \tag{C.77}$$

The following results will be used to prove the above statements:

$$Y > 1 > Z \tag{C.78}$$

$$\Rightarrow \left(\frac{1}{Z}\right) > \left(\frac{1}{Y}\right) > 0 \tag{C.79}$$

$$\Rightarrow 1 > \left(\frac{Y}{Y+Z}\right) > \left(\frac{Z}{Y+Z}\right) > 0 \tag{C.80}$$

$$\Rightarrow 0 > \ln\left(\frac{Y}{Y+Z}\right) > \ln\left(\frac{Z}{Y+Z}\right) \tag{C.81}$$

$$\Rightarrow -\ln\left(\frac{Z}{Y+Z}\right) > -\ln\left(\frac{Y}{Y+Z}\right) > 0 \tag{C.82}$$

Statement 1: *First Transition Configuration Entropy Contribution*

$$-\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{1}{2}\right) > -\left(\frac{Z}{Y+Z}\right)\ln\left(\frac{1}{2}\right)$$

It has already been shown that:

$$1 > \left(\frac{Y}{Y+Z}\right) > \left(\frac{Z}{Y+Z}\right) > 0$$

Since:

$$\ln\left(\frac{1}{2}\right) < 0$$

$$-\ln\left(\frac{1}{2}\right) > 0$$

By the multiplicative property of inequalities:

$$a > b > 0$$

$$c > 0$$

$$\Rightarrow ac > bc$$

Therefore:

$$-\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{1}{2}\right) > -\left(\frac{Z}{Y+Z}\right)\ln\left(\frac{1}{2}\right)$$

Q.E.D.

Statement 2: *Transition Entropy Contribution*

$$\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{X}{Z}\right)+\left(\frac{Y}{Y+Z}\right)p\ln\left(\frac{Z}{Z}\right)>\left(\frac{Z}{Y+Z}\right)\ln\left(\frac{X}{Y}\right)+\left(\frac{Z}{Y+Z}\right)p\ln\left(\frac{Y}{Z}\right)$$

Expanding terms and simplifying gives:

$$\left(\frac{Y}{Y+Z}\right)\ln(X)+\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{1}{Z}\right)>\left(\frac{Z}{Y+Z}\right)\ln(X)+\left(\frac{Z}{Y+Z}\right)\ln\left(\frac{1}{Y}\right)+\left(\frac{Z}{Y+Z}\right)p\ln(Y)+\left(\frac{Z}{Y+Z}\right)p\ln\left(\frac{1}{Z}\right)$$

It has already been shown that:

$$1>\left(\frac{Y}{Y+Z}\right)>\left(\frac{Z}{Y+Z}\right)>0$$

Since Z is sub-unity while X and Y are not:

$$\ln\left(\frac{1}{Z}\right)>0$$

$$\ln(X)>0$$

By the multiplicative property of inequalities:

$$a>b>0$$

$$c>0$$

$$\Rightarrow ac>bc$$

Thus, it is clear the following contributions are true:

$$\left(\frac{Y}{Y+Z}\right)\ln(X)>\left(\frac{Z}{Y+Z}\right)\ln(X)$$

$$\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{1}{Z}\right)>\left(\frac{Z}{Y+Z}\right)\ln\left(\frac{1}{Z}\right)$$

Thus, it is necessary to show:

$$0 \geq \left(\frac{Z}{Y+Z}\right) \ln\left(\frac{1}{Y}\right) + \left(\frac{Z}{Y+Z}\right) p \ln(Y)$$

Rearranging terms on the right hand side:

$$0 \geq \left(\frac{Z}{Y+Z}\right) \ln\left(\frac{1}{Y}\right) (1-p)$$

The following components are positive:

$$(1-p) \geq 0$$

$$1 > \left(\frac{Z}{Y+Z}\right) > 0$$

However, there is exactly one component that is always negative:

$$Y > 1$$

$$\Rightarrow 0 > \ln\left(\frac{1}{Y}\right)$$

Multiplying the components out guarantees a negative term for p less than 1. If p = 1, then the equality holds. Thus,

$$0 \geq \left(\frac{Z}{Y+Z}\right) \ln\left(\frac{1}{Y}\right) (1-p)$$

Q.E.D.

Statement 3: *First Transition New State Entropy Contribution*

$$\left(\frac{Y}{Y+Z}\right)\ln\left(\frac{1}{Z}\right) > \left(\frac{Z}{Y+Z}\right)\ln\left(\frac{1}{Y}\right)$$

Since Z is sub-unity and Y is not:

$$\ln\left(\frac{1}{Z}\right) > 0$$

$$\ln\left(\frac{1}{Y}\right) < 0$$

Furthermore, it has already been shown that:

$$1 > \left(\frac{Y}{Y+Z}\right) > 0$$

$$1 > \left(\frac{Z}{Y+Z}\right) > 0$$

The left-hand side is always positive and the right-hand side is always negative. Therefore, the inequality must always be true.

Q.E.D.

Statement 4: *Second Transition Configuration Entropy Contribution*

$$-\left(\frac{Y}{Y+Z}\right) p \ln p' > -\left(\frac{Z}{Y+Z}\right) p \ln p'$$

Multiplying through by -1:

$$\left(\frac{Y}{Y+Z}\right) p \ln\left(\frac{1}{p'}\right) > \left(\frac{Z}{Y+Z}\right) p \ln\left(\frac{1}{p'}\right)$$

Since p' satisfies the conditions of a probability expression:

$$0 < p' < 1$$

Rearranging terms gives:

$$\left(\frac{1}{p'}\right) > 1$$

$$\ln\left(\frac{1}{p'}\right) > 0$$

p also satisfies the conditions of a probability expression:

$$0 < p < 1$$

Thus,

$$p \ln\left(\frac{1}{p'}\right) > 0$$

It has already been shown that:

$$1 > \left(\frac{Y}{Y+Z}\right) > \left(\frac{Z}{Y+Z}\right) > 0$$

By the multiplicative property of inequalities:

$$\left(\frac{Y}{Y+Z}\right) p \ln\left(\frac{1}{p'}\right) > \left(\frac{Z}{Y+Z}\right) p \ln\left(\frac{1}{p'}\right)$$

Q.E.D.

Statement 5: *Second Transition New State Entropy Contribution*

$$\left(\frac{Y}{Y+Z}\right) p \ln\left(\frac{1}{Z}\right) > \left(\frac{Z}{Y+Z}\right) p \ln\left(\frac{1}{Z}\right)$$

Since Z is sub-unity:

$$\ln\left(\frac{1}{Z}\right) > 0$$

and since p satisfies the condition of a probability expression:

$$1 > p > 0$$

It is clear that:

$$p \ln\left(\frac{1}{Z}\right) > 0$$

It has already been shown that:

$$1 > \left(\frac{Y}{Y+Z}\right) > \left(\frac{Z}{Y+Z}\right) > 0$$

Thus, by the multiplicative property of inequalities:

$$\left(\frac{Y}{Y+Z}\right) p \ln\left(\frac{1}{Z}\right) > \left(\frac{Z}{Y+Z}\right) p \ln\left(\frac{1}{Z}\right)$$

Q.E.D.

Statement 6: *Future Entropy Contribution*

$$\left(\frac{Y}{Y+Z}\right)pH > \left(\frac{Z}{Y+Z}\right)pH$$

By the theorem condition:

$$H > 0$$

Since p satisfies the condition of a probability expression:

$$1 > p > 0$$

It has already been shown that:

$$1 > \left(\frac{Y}{Y+Z}\right) > \left(\frac{Z}{Y+Z}\right) > 0$$

Thus, by the multiplicative property of inequalities:

$$\left(\frac{Y}{Y+Z}\right)pH > \left(\frac{Z}{Y+Z}\right)pH$$

Q.E.D.

Appendix D

Network Application, Traffic Behavior, and Quality of Service Table

Network Application, Traffic Behavior, and Quality of Service. Information collated from Cisco, Chen-Nee Chuah.[91,92]

Service Type	Example Applications	Traffic Behavior	QoS Description	Data Rate	Loss	Round-trip Latency	Round-trip Jitter
VoIP		Constant bit rate, or bursty traffic (with silence suppression).	Very sensitive to delay and jitter. Sensitive to loss. Bandwidth requirement: Low and predictable. Requires predictable delay and loss.	21-320 kbps	1 %	0.3 sec	0.06 sec
Streaming Video	Television, Youtube			Depends	5 %	10 sec	-
Interactive Video	Video conferencing	Constant or variable bit rate (depending on the codec).	Extremely sensitive to delay and jitter, and loss. Bandwidth requirement: High. Requires predictable delay and loss.	384 kbps	1 %	0.3 sec	0.06 sec
Interactive Data	Telnet, Oracle Thin-Clients, AOL Instant Messenger, Yahoo! Instant Messenger, PlaceWare (Conference), Netmeeting Whiteboard.	Highly interactive applications with tight user-feedback requirements.	Very sensitive.				
Transactional Data	SAP, PeopleSoft-Vantive, Oracle-Financials, Internet Procurement, B2B, Supply Chain Management, Application Server, Oracle 8i Database, Microsoft SQL.	Transactional applications typically use a client/server protocol model. User-initiated, client-based queries are followed by server response. The query response can consist of many messages between client and server. The query response can consist of many TCP and FTP sessions running simultaneously (for example, HTTP-based applications). Many small two-way transactions, 'chatty'.	Sensitive to loss and delay. Bandwidth requirement: Low to moderate. Best effort, must be stable and reliable.				
Bulk Data	Database syncs, network-based backups, Lotus Notes, Microsoft Outlook, e-mail download (SMTP, POP3, IMAP, Exchange), video content distribution, large FTP file transfers.	Long file transfers. Always invokes TCP congestion management. The Bulk Data class is intended for applications that are relatively noninteractive and not drop sensitive, and that typically span their operations over a long period of time as background occurrences.	Very tolerant of delay and loss. Bandwidth requirement: Low. Moderate bandwidth.				
Best-effort Data	All noncritical traffic, HTTP web browsing, other miscellaneous traffic.	A series of small, bursty file transfers.	Tolerant of moderate delay and loss. Bandwidth requirement: Low to moderate.				

References

[1] Douglas J. Mudgway. Uplink-Downlink: A History of the Deep Space Network, 1957-1997. NASA History Series, 2001. SP-2001-4227.

[2] Frank L. Lambert. Obstructions to the second law make life possible. http://www.2ndlaw.com/obstructions.html, February 2008.

[3] Will Ivancic. Internet protocol version 6 and network centric operations key concepts. http://roland.grc.nasa.gov/~ivancic/papers_presentations/2007/Understanding_IPv6.ppt, 2004. Original source: PC of Japan.

[4] Pok-Ching Lee. Recent works. http://www.ccs.neu.edu/home/lpok/, 2002.

[5] Wikipedia. Activation energy. http://en.wikipedia.org/wiki/Activation_energy, December 2008.

[6] Olivier L. de Weck, Uriel Scialom, and Afreen Siddiqi. Optimal reconfiguration of satellite constellations with the auction algorithm. Acta Astronautica, 2007. doi: 10.1016/j.actaastro.2007.02.008.

[7] Jennifer E. Underwood. Distributed satellite communications system design: first-order interactions between system and network architectures. Master's thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 2005.

[8] S. Chan, A. Samuels, N. Shah, J. Underwood, and O.L. de Weck. Optimization of hybrid satellite constellations using multiple layers and mixed circular-elliptical orbits. In 22nd AIAA International Communications Satellite Systems Conference and Exhibit, Monterey, California, May 9-12 2004. AIAA. AIAA-2004-3205.

[9] Satellite Standard Group. Subscriber breakdowns. http://satellitestandard.blogspot.com/2007/01/subscriber-beakdowns.html, January 2007.

[10] Wikipedia. Xm satellite radio. http://en.wikipedia.org/wiki/XM_Satellite_Radio, December 2008.

[11] NASA. Exploration systems architecture study. http://www.nasa.gov/pdf/140649main_ESAS_full.pdf, November 2005. NASA-TM-2005-214062.

[12] B. Bairstow, M. Baldesarra, T. Coffee, A. Fong, W.K. Hofstetter, S.B. Hong, D. Judnick, S. McCloskey, J. Mellein, and J. Underwood. Extensible planetary surface mobility systems, final report. Technical report, Massachusetts Institute of Technology, Spring 2006. 16.89 / ESD 352 Space Systems Engineering course.

[13] Robin Cowan. Nuclear power reactors: A study in technological lock-in. http://www.econ.nyu.edu/cvstarr/working/1988/RR88-33.pdf, October 1988.

[14] Jim McGee. Legacy systems: Why history matters. http://esj.com/news/article.aspx?EditorialsID=1529, October 2005.

[15] Matthew Silver. Time-expanded decision networks: A framework for designing evolvable complex systems. Systems Engineering, 10(2):167–186, 2007.

[16] Matheiu Chaize. Enhancing the economics of satellite constellations via staged deployment and orbital reconfiguration. Master's thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 2003.

[17] KnowledgeLeader. Legacy value engineering: The fountain of youth to maximize the value of existing technology investments. http://www.knowledgeleader.com/KnowledgeLeader/Content.nsf/Web+Content/WhitePapersArticlesLegacyValueEngineering!OpenDocument, April 2003.

[18] Arizona State University Design Automation Laboratory. Legacy systems engineering: Background. http://asudesign.eas.asu.edu/projects/cmminspres.html, 2007.

[19] Fred Nickols. Strategy: Definitions and meaning. http://home.att.net/~nickols/strategy_definition.htm, 2000.

[20] Olivier L. de Weck. M.i.t. strategic engineering: Designing systems for an uncertain future. http://strategic.mit.edu/, December 2008.

[21] Dictionary.com L.L.C. Dictionary.com: Advantage. http://dictionary.reference.com/browse/Advantage, December 2008.

[22] Dictionary.com. Value. http://dictionary.reference.com/browse/value, January 2009. Source: Princeton University.

[23] D.K. Hitchins. Advanced Systems Thinking, Engineering, and Management. Artech House, Boston, MA, 2003.

[24] Ernst Fricke and Armin P. Schulz. Design for changeability (dfc): Principles to enable changes in systems throughout their entire lifecycle. Systems Engineering, 8(4):1098–1241, 2005.

[25] Afreen Siddiqi. Reconfigurability in space systems: architecting framework and case studies. PhD thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 2006.

[26] III John A. Christian and John R. Olds. A quantitative methodology for identifying evolvable space systems. In 1st Space Exploration Conference: Continuing the Voyage of Discovery, Orlando, Florida, 30 January - 1 February 2005. AIAA-2005-2543.

[27] G.A. Bounova and O.L. de Weck. Graph-theoretical considerations in design of large telescope arrays for robustness and scalability. In 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Austin, Texas, 18 - 21 April 2005. AIAA-2005-2063.

[28] Marie L. Garcia and Olin H. Bray. Fundamentals of technology roadmapping. http://www.sandia.gov/PHMCOE/pdf/Sandia'sFundamentalsofTech.pdf, April 1997.

[29] Carliss Y. Baldwin, editor. Design rules: the power of modularity, volume 1. MIT Press, 2000.

[30] F.A. Ohemeng and T. Mole. Value-focused approach to built asset renewal and replacement decision. In Proceedings of the Royal Institute of Chartered Surveyors COBRA Conference, London, 1996. RICS.

[31] Marguerite Reardon. IPv6 coming to america. http://www.lightreading.com/document.asp?doc_id=36168, June 2003.

[32] Microsoft Corporation. IPv6 for microsoft windows: Frequently asked questions. http://www.microsoft.com/technet/network/ipv6/ipv6faq.mspx, September 2002.

[33] Geoff Huston. Ipv4 address space report. http://www.potaroo.net/tools/ipv4/index.html, January 2009.

[34] Tony Hain. A pragmatic report on ipv4 address space consumption. http://www.tndh.net/~tony/ietf/ipv4-pool-combined-view.pdf, July 2008.

[35] Mike Chapple. Is a transition from IPv4 to IPv6 worth the effort? http://searchsecurity.techtarget.com/expert/KnowledgebaseAnswer/0,289625,sid14_gci1252252,00.html, February 2007.

[36] Charles M. Kozierok. The TCP/IP guide: A TCP/IP reference you can understand! http://www.tcpipguide.com/free/t_InternetProtocolVersion6IPv6IPNextGenerationIPng.htm, September 2005.

[37] G. Pujolle, editor. Management, Control and Evolution of IP Networks. International Scientific and Technical Encyclopedia, 2007.

[38] National Oceanic and Atmospheric Administration (NOAA). Rockets, radar, and computers: The international geophysical year. http://celebrating200years.noaa.gov/magazine/igy/welcome.html#intro, May 2007.

[39] Constance McLaughlin Green and Milton Lomask. Vanguard: A history. http://history.nasa.gov/SP-4202/chapter1.html, 1970. NASA SP-4202.

[40] The National Academies (NAS). The international geophysical year. http://www.nas.edu/history/igy/, 2005.

[41] Andrew J. LePage. Vanguard: America's answer to sputnik. http://seds.org/spaceviews/9712/articles3.html, December 1997.

[42] Peter Byrd. Political dictionary: Cold war. http://www.answers.com/topic/cold-war, Viewed February 2007.

[43] Roger D. Launius. Sputnik and the origins of the space age. http://history.nasa.gov/sputnik/sputorig.html.

[44] Homer Edward Newell. Beyond the Atmosphere: Early Years of Space Science. Scientific and Technical Information Branch, National Aeronautics and Space Administration, 1980.

[45] Mark Wade. Sputnik plus 50. http://www.astronautix.com/articles/spulus50.htm, 1997 - 2008.

[46] Steve Garber. Sputnik and the dawn of the space age. http://history.nasa.gov/sputnik/, October 2007.

[47] Craig B. Watt. The road to the deep space network. IEEE Spectrum, 30(4), April 1993.

[48] Douglas S. Abraham. Future mission trends and their implications for the deep space network. In Space 2006, San Jose, California, 19 - 21 September 2006. AIAA 2006-7247.

[49] Dictionary.com. Framework. http://dictionary.reference.com/browse/framework, January 2009. Source: The American Heritage Dictionary of the English Language, Fourth Edition.

[50] Stephen Lawson. Iridium plans next-generation, ip-based satellites. http://www.pcworld.com/article/129151/iridium_plans_nextgeneration_ipbased_satellites.html, February 2007.

[51] Ernst Fricke, Bernd Gebhard, Herbert Negele, and Eduard Igenbergs. Coping with changes: Causes, findings, and strategies. Systems Engineering, 3(4):169–179, 2000.

[52] J. Etkin, J.A. Zinky, and J. Papadopoulos. Crises in network evolution: three case studies and one proposed solution. In Intelligent Networks and their Applications, Zurich Switzerland, 16-19 March 1992. 1992 International Zurich Seminar on Digital Communications.

[53] INCOSE. Systems Engineering Handbook. INCOSE, version 3 edition, 2004.

[54] Wiley J. Larson and James R. Wertz, editors. Space Mission Analysis and Design. Microcosm Press, 3 edition, October 1999.

[55] Willard L. Simmons. A framework for decision support in systems architecting. PhD thesis, Massachusetts Institute of Technology, 2008. Department of Aeronautics and Astronautics.

[56] Lunar architecture update. http://www.nasa.gov/pdf/190083main_AIAA_ESMD_finalSPACE2007.pdf, September 2007. Presentation at AIAA Space 2007.

[57] Wilfried K. Hofstetter, Paul D.Wooster, and Edward F. Crawley. Analysis of human lunar outpost strategies and architectures. In Space 2007, Long Beach, CA, USA, September 2007. AIAA.

[58] Adam M. Ross and Daniel E. Hastings. Assessing changeability in aerospace systems architecting and design using dynamic multi-attribute tradespace exploration. In Space 2006, San Jose, California, September 19-21 2006. AIAA. AIAA 2006-7255.

[59] PC of Japan. Transition and operations costs. `roland.grc.nasa.gov/~ivancic/papers_presentations/2007/Understanding_IPv6.ppt`.

[60] Richard de Neufville. Applied Systems Analysis: Engineering Planning and Technology Management. McGraw-Hill, Inc., New York, New York, USA, 1990.

[61] Wikipedia. Entropy. `http://en.wikipedia.org/wiki/Entropy`, December 2008.

[62] F. Heylighen. Entropy increase: the box with two compartments. `http://pespmc1.vub.ac.be/ENTROBOX.html`, May 2003.

[63] Frank L. Lambert. Entropy and the second law of thermodynamics. `http://www.2ndlaw.com/entropy.html`, February 2008.

[64] Wikipedia. Configuration entropy. `http://en.wikipedia.org/wiki/Configuration_entropy`, September 2008.

[65] Wikipedia. Entropy (information theory). `http://en.wikipedia.org/wiki/Information_entropy`, December 2008.

[66] Inder Jeet Taneja. Properties of shannon's entropy. `http://www.mtm.ufsc.br/~taneja/book/node6.html`, June 2001.

[67] Inder Jeet Taneja. Measures of uncertainty: Shannon's entropy. `http://www.mtm.ufsc.br/~taneja/book/node5.html`, June 2001.

[68] Peter Fleissner and Wolfgang Hofkirchner. Entropy and its implications for sustainability. `http://http://igw.tuwien.ac.at/igw/Menschen/hofkirchner/papers/InfoConcept/Entropy/entropy.htm`. In: Dragan, J.C., Demetrescu, M.C., Seifert, E.K.

(Eds.), Implications and Applications of Bioeconomics, Proceedings of the Second International Conference of the E.A.B.S., Palma de Mallorca, 1994, Edizioni Nagard, Milano 1997, 147-155.

[69] Wikipedia. Hilbert space. `http://en.wikipedia.org/wiki/Hilbert_space`, December 2008.

[70] M. Mohorcic, M. Werner, A. Svigelj, and G. Kandus. Comparison of adaptive routing algorithms in isl networks considering various traffic scenarios. 4th European Workshop on Mobile and Personal Communications, September 2000.

[71] T. Kashitani. Development and application of an analysis methodology for satellite broadband network architectures. In 20th AIAA ICSSC and Exhibit, Montreal, Canada, May 12-15 2002. AIAA.

[72] D. Chang. Quantifying technology infusion and policy impact on low earth orbit communication satellite constellations. Master's thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 2004.

[73] Satellite telecommunications: Market perspectives and industrial situation. Technical report, ESA Telecommunications, 2005.

[74] Olivier de Weck, Richard de Neufville, and Mathieu Chaize. Staged deployment of communications satellite constellations in low earth orbit. Journal of Aerospace Computing, Information, and Communication, 1(3), March 2004.

[75] D.D. Chang and O. de Weck. Basic capacity calculation methods and benchmarking for mf-tdma and mf-cdma communication satellites. International Journal of Satellite Communications and Networking, 23(3), May/June 2005.

[76] Lloyd Wood. Lloyd's satellite constellations: Iridium. `http://personal.ee.surrey.ac.uk/Personal/L.Wood/constellations/iridium.html`, May 2006.

[77] O. L. de Weck and D. Chang. Architecture trade methodology for leo personal communication systems. In 20th International Communications Satellite Systems Conference, Montréal, Québec, Canada, May 2002. AIAA. AIAA-2002-1866.

[78] Satellite telephone quality of service comparison: Iridium vs. globalstar. Technical report, Frost & Sullivan, 2002.

[79] Margaret M. McMahon and Robert Rathburn. Measuring latency in iridium satellite constellation data services. http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/233.pdf, June 2005. Defense Technical Information Center (DTIC).

[80] NASA. The vision for space exploration. http://history.nasa.gov/Vision_For_Space_Exploration.pdf, February 2004. Document outlining the plan given by President George W. Bush in a speech.

[81] Draper and MIT. From value to architecture: The exploration system of systems. http://www.nasa.gov/zip/187555main_cer_final.zip, March 2005. NASA Concept Exploration and Refinement Study Phase I Final Presentation.

[82] Christopher E. Carr. Distributed architectures for mars surface exploration. Master's thesis, Massachusetts Institute of Technology, Department of Aeronautics and Astronautics, 2001.

[83] Apollo lunar surface journal. http://www.hq.nasa.gov/office/pao/History/alsj/frame.html, October 2008.

[84] Trudy E. Bell. A new paradigm for lunar orbits. http://science.nasa.gov/headlines/y2006/30nov_highorbit.htm, November 2006.

[85] Todd A. Ely and Erica Lieb. Constellations of elliptical inclined lunar orbits providing polar and global coverage. In Astrodynamics Specialists Conference, Lake Tahoe, CA, August 7-11 2005. AAS/AIAA. AAS 05-343.

[86] J.D. Parsons. The Mobile Radio Propagation Channel. John Wiley & Sons, LTD, New York, New York, USA, 2 edition, 2000.

[87] Konstantinos Kalligeros. Using real options for engineering systems design framework, applicability and limitations. In Engineering Systems Symposium. MIT, March 29-31 2004.

[88] Wikipedia. Decision theory. http://en.wikipedia.org/wiki/Decision_theory, January 2009.

[89] Ann Svendsen and Myriam Laberge. Mapping a stakeholder network. Technical report, CoreRelation Consulting Inc., April 2007.

[90] McGraw-Hill. Dictionary of scientific and technical terms, October 2002. Radial and Hydrostatic Bearing definitions, via answers.com.

[91] Tim Szigeti and Christina Hattingh. Quality of service design overview. `http://www.ciscopress.com/articles/article.asp?p=357102&seqNum=3`, December 2004.

[92] Chen-Nee Chuah. Providing end-to-end qos for ip-based latency-sensitive applications. `http://www.ece.ucdavis.edu/~chuah/paper/proposal.pdf`, October 1999.