AIAA SPACE 2011 Conference & Exposition
27 - 29 September 2011, Long Beach, California

AIAA 2011-7274

A Markovian State-Space Flexibility Framework Applied to Distributed-Payload Satellite Design Decisions

Jarret M. Lafleur*
Georgia Institute of Technology, Atlanta, Georgia 30332

Over the past decade, the space industry has increasingly recognized the need for new systems to be designed for flexibility, or the capability to be easily modified in response to changes in future requirements or environments. Despite widespread interest, however, the state of the art in designing flexibility into space systems today remains limited. To address these limitations, this paper presents the basis of a quantitative, stochastic, multi-objective, and multi-period framework for integrating flexibility into space system design decisions. Central to the framework are five steps that (1) define configuration options and transition costs, (2) define a stochastic model for mission demand environment changes, (3) link configurations and demand environments via quantitative performance metrics, (4) identify Pareto-optimal configuration paths and decision policies, taking advantage of efficient multi-objective Markov decision process techniques, and (5) utilize these path and policy results to inform initial system selection. The framework is applied to a realistic example in which design decisions are suggested for a hypothetical multi- or distributed-payload satellite system. The application illustrates how flexibility-informed trades can permit selection of a satellite system that most effectively responds to uncertain future demands.

Nomenclature

A	=	set of available actions or decisions	s	=	particular total state
a	=	particular action or decision	T	=	total number of time periods in time horizon
b	=	per-period budget level	t	=	current time period
C	=	total cost transition matrix	u	=	system utility
$c_{i,j}$	=	element of total cost transition matrix	w_i	=	weight on the i^{th} objective
C_{dev}	=	development cost transition matrix	y_i	=	per-period i^{th} objective performance
C_{rec}	=	recurring cost transition matrix	α	=	transition cost budget threshold
G	=	set of available next-period positions	β	=	multiplicative discounting factor
g	=	set of zero-cost next-period positions	γ	=	transition cost
h	=	aggregate per-period objective function	ι	=	reference system index
J	=	cumulative aggregate objective function	κ	=	reference time period index
Q	=	demand environment state random variable	λ	=	next-period system index
M	=	number of objectives	σ	=	current operating/demand environment state
N	=	number of candidate next-period systems	ς	=	prior operating/demand environment state
n	=	objective function power	Φ_i	=	number of transitions available from Config. i
p	=	conditional state transition probability	χ	=	starting position
S	=	set of all total states	ψ	=	next-step position

I. Introduction

OVER the past decade, the Department of Defense, National Aeronautics and Space Administration (NASA), and other organizations with substantial stake in space systems have increasingly recognized the need for new spacecraft and space architectures to be designed for flexibility, or the capability to be easily modified in response to environments or requirements that may materialize months or years after these systems are fielded. Recent high-profile examples include the Defense Advanced Research Projects Agency (DARPA) System F6 program[1-3], which seeks to demonstrate the flexibility of a wirelessly-connected cluster of free-flying satellites, as well as the U.S.

*Ph.D. Candidate, Daniel Guggenheim School of Aerospace Engineering, Student Member AIAA.

Human Spaceflight Plans (Augustine) Committee's "flexible path" option for the future of U.S. human spaceflight[4], which seeks to enable human mission options to a variety of inner solar system destinations.

In general, this desire for flexibility is driven by an acknowledgement that future demands and expectations on a space system cannot be predicted with certainty – and a realization that accounting for this fact during early system design phases could result in important design decisions that might otherwise be overlooked. Ultimately, flexibility could prove critical to a system's ability to effectively respond to uncertain future events, whether these events take the form of mission-jeopardizing risks (e.g., for satellites, physical or directed energy attack, program funding cuts, or decreases in user demand for services) or mission-enhancing opportunities (e.g., increases in user demand or program funding). However, posing this problem quantitatively is challenging. Today, the objective of flexibility is frequently considered only qualitatively during design and system selection processes. Occasionally, quantitative analysis is conducted but tends to be deterministic, single-objective, and/or limited to consider only one future time period. In contrast, most practical system design problems are non-deterministic, concerned with decisions among competing objectives, and concerned with performance over multiple future time periods. The question remains: How can space systems engineers and decision-makers systematically, quantitatively, objectively, and pragmatically consider flexibility in the design of a new space system?

To contribute a further step toward development in this area, this paper presents the basis of a quantitative, stochastic, multi-objective, and multi-period framework for integrating flexibility into space system design decisions. Central to the framework are five steps that model the systems and decisions of interest and subsequently provide multi-period and multi-objective decision support. The framework itself draws from literature and tools within the fields of industrial engineering, aerospace engineering, and economics in order to operationally define flexibility and transform its consideration into a tractable problem of stochastic optimal control.

This paper is organized as follows: Section II provides highlights from the academic literature on flexibility and state of the practice in designing flexibility into aerospace systems. Section III introduces this paper's proposed framework, and Section IV illustrates the framework's application in the context of a distributed-payload defense satellite system design decision. Section V contributes a concluding discussion.

II. Flexibility Literature Highlights and State of the Practice

The Merriam-Webster Dictionary defines flexibility as the "ready capability to adapt to new, different, or changing requirements."[5] This paper adopts a similar definition, namely that flexibility is *the capability to easily modify a system after it has been fielded in response to a changing environment or changing requirements* (cf. Ref. 6). Central to this notion of flexibility are the conditions that (1) a system's environment or requirements may change in the future and (2) the system can, to some degree, be modified to accommodate such change. This definition also includes the notion of ease of modification: All else being equal, one system is more flexible than another if it takes less effort or fewer resources to accomplish the same change. These basic concepts form a common thread within the past century's history of thought on flexibility; however, only within the past decade has the aerospace engineering community begun to develop quantitative methods for considering this property in the context of system design decisions. This section surveys from past literature on flexibility, highlighting important concepts from the fields of economics, industrial engineering, and aerospace engineering.

A. Early Economic Notions of Flexibility

Some of the earliest discussions on flexibility in a decision-making context originate in the economics literature. As early as 1921, economist Frank Knight observed that, compared to agricultural production, which requires commitment at the beginning of each growing season, the supply of manufactured goods "is more flexible over short periods of time" since these goods can be stored and the decision about whether to bring them to the market can be delayed.[7] Sixteen years later, Hart recognized that the postponement of decisions is a normal occurrence and preserves flexibility in a business plan.[8] However, he also recognized that this flexibility generally comes at a cost:

> The entrepreneur's fundamental means of meeting uncertainty is the postponement of decisions till more information comes in – that is to say, the preservation of *flexibility* in his business plan. But flexibility involves costs … ordinarily a given production-schedule can be produced at lower cost if the entrepreneur has adapted his input to it well in advance than if plans are improvised.[8]

In 1939, Stigler developed economic thought on flexibility somewhat further. He too recognized that "flexibility will not be a 'free good'"[9] but also illustrated how a flexible plant might have a smaller variability in average and marginal costs as a function of output compared to an inflexible plant.

In 1964, Koopmans reiterated the relevance of flexibility by observing that "almost all choices occurring in real life are sequential, 'piece-meal,' choices between alternative ways of narrowing down the presently existing opportunity rather than 'once-and-for-all' choices between specific programs visualized in full detail."[10] Koopmans introduced the notion of "partitioning of opportunities" which, as shown in Fig. 1, modeled the narrowing of opportunities with time as a tree of opportunity nodes spaced at discrete times in the future. Koopmans' partitioning of opportunities resembles decision tree analysis, introduced in the late 1950s and 1960s within the broader field of decision analysis.[11-17] Decision tree analysis has been used substantially in management, economics, and engineering contexts (for examples, see Refs. 16-21), typically for the cases in which a user's objective is minimization or maximization of the expected value of a single profit, cost, or utility metric. A common drawback of the approach is that the analysis (and even simply populating the tree's probability inputs) can quickly become unwieldy as the number of options and time periods grow into a "decision bush" rather than a "decision tree".[20-21] Also, typically the focus of decision tree analysis is on valuating existing options rather than recommending which options should be embedded into the system initially.[22] Nevertheless, recognition that the options provided by flexibility can be visualized in a rapidly-expanding tree structure provides a highly useful model for discussion and thought. It also hints that dynamic programming techniques, which are well-suited to optimizing paths within networks of nodes, may be particularly useful in analysis of flexibility.

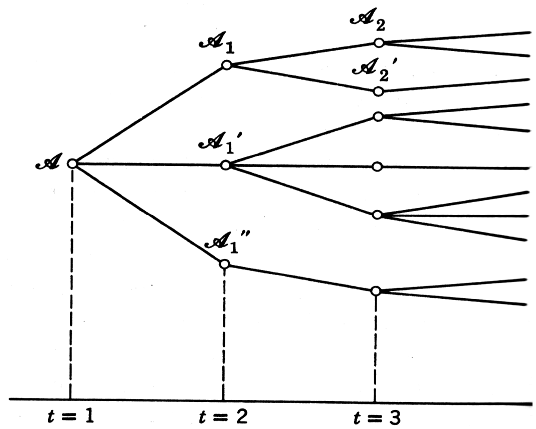

Figure 1. Visualization of Koopmans' partitioning of opportunities[10].

B. The Two-Period State-Centric Notion of Flexibility

A second and largely separate body of literature in economics and industrial engineering considers flexibility within a framework of period-to-period transitions between options in a state-space. Epitomizing this view is a paper written in 1984 by Jones and Ostroy[23] which suggested, "Flexibility is a property of initial positions. It refers to the cost, or possibility, of moving to various second period positions." The authors also suggested, "One position is more flexible than another if it leaves available a larger set of future positions at any given level of cost." This was mathematically formalized with Eqs. (1) and (2). Eq. (1) defines $G(\chi,\sigma,\alpha)$ as the set of next-period positions ψ attainable from position χ at a cost γ that does not exceed some value α, in the context of some state σ of the operating environment. Eq. (2) formalizes that position χ is more flexible than χ' (denoted by $\chi >_F \chi'$) if the set of positions attainable from χ always contains the set attainable from χ', excluding the zero-cost option to stay in χ'.

$$G(\chi,\sigma,\alpha) \equiv \{\psi : \gamma(\chi,\psi,\sigma) \leq \alpha\} \tag{1}$$

$$\begin{aligned} &\chi >_F \chi' \quad \text{when} \\ &G(\chi,\sigma,\alpha) \supset G(\chi',\sigma,\alpha) \setminus g(\chi') \end{aligned} \tag{2}$$

Thus, an important recognition in Jones and Ostroy's work is that the relative flexibility of two positions is budget-dependent (or resource-dependent). For an infinite budget, two positions would be equally flexible because each can reach the same set of [all possible] future positions. At lower budgets, this may not be true. However, Eq. (2) has a limitation: It defines relative flexibility only for the case where the set of second-period positions from χ' is fully contained within the set of second-period positions from χ. No conclusion can be drawn if one of the sets is not fully contained within the other. This is appropriate in principle, as the positions available from χ' that are not available from χ may be very important (e.g., may perform particularly well in meeting a particular new requirement or environment), and it illustrates the need to consider more than cost when making decisions regarding flexibility.

Other works which have advanced similar state-centric frameworks for considering flexibility include Christian and Olds,[24,25] Gupta and Rosenhead,[26] Baykasoğlu[27], Silver and de Weck[28-29], and Mandelbaum and Buzacott[30]. In general, these frameworks and others of this class are helpful because they provide a visualization of the concept of flexibility itself (as opposed to the value of flexibility), which is intuitively related to the number of options that exist for a system over time. However, these frameworks tend to be difficult to visualize and apply for decisions consisting of more than two periods.

C. Flexibility in Aerospace System Design Literature

Recently, the Department of Defense, NASA, and other aerospace organizations have increasingly emphasized the need for new aerospace systems to be designed for flexibility. High-profile examples such as DARPA's System F6 program and the Augustine Committee's "flexible path" option for human space exploration have sparked a number of studies within the aerospace academic literature.

Ross, Viscito, and Rhodes[31-32] propose the analysis of flexibility in terms of epochs and eras, where an epoch is a time period of "fixed context and fixed value expectations"[32] and an era is a time-ordered sequence of epochs. Once an era is defined, Ross and Viscito[31] propose quantifying flexibility via a metric called value-weighted filtered outdegree (VWFO) as defined in Eq. (3). In this equation, $u_\lambda^{\kappa+1}$ indicates the utility of system design option λ in epoch $\kappa+1$, and $Arc_{\iota,\lambda}^{\kappa}$ is a binary 0 or 1 depending on whether the transition is possible for a given budget. As a result, systems with many high-utility next-epoch (next-period) options and few low-utility next-epoch options receive high VWFO scores. However, this metric has some limitations. First, the use of the signum function in the summation of Eq. (3) permits a system with many high-utility options and correspondingly many low-utility options to have a VWFO indistinguishable from one with only medium-utility options. Second, VWFO is computed from epoch to epoch, making it difficult to assess for an entire era. Finally, the metric convolves the flexibility with the value (or utility) of that flexibility, preventing the two from being distinguished. However, the metric contributes a clear example employing a two-period state-centric concept of flexibility, including use of a budget constraint.

$$VWFO_\iota^\kappa = \frac{1}{N-1}\sum_{\lambda=1}^{N-1}\left[\operatorname{sgn}\left(u_\lambda^{\kappa+1} - u_\iota^{\kappa+1}\right) * Arc_{\iota,\lambda}^{\kappa}\right] \tag{3}$$

Greater depth on the flexibility problem was covered in theses by Saleh[21] and later Mark[33] and Nilchiani[34]. In 2002, Saleh[21] extensively motivated the need for flexibility in space systems and examined its definition, in particular contrasting it against the more static property of robustness. Specific examples were provided to illustrate the need for flexibility in modern space systems, including instances of historical requirements change, market demand change, and obsolescence. Saleh applied techniques from decision tree and real options analysis to demonstrate the existence of net-present-value-optimal design lifetimes for revenue-generating satellites and used these techniques further to quantify the value of satellite servicing.

In 2005, Mark[33] further explored flexibility for the example of an unmanned aerial vehicle. Mark proposed considering flexibility in the context of platforms and frames, where a platform is the set of common elements between modified designs and a frame is a set of changed elements. Mark proposed to define flexibility as "the ratio of performance enhancement (output) to the cost and time required to realize such an enhancement (inputs)"[33]. Later in 2005, Nilchiani[34] proposed a 12-step process for assessing the value of flexibility in a space system, which included using decision trees as well as creating a "flexibility tradespace" for visualizing alternatives' cost-revenue (and/or cost-benefit) trades one period into the future. Nilchiani also addressed how the proposed methodology could be integrated into a multi-attribute trade-space exploration in a methodology named FlexiMATE.

In 2009, Lim[35-36] also proposed a general approach to design evolution, focusing on aircraft and using example applications of evolving the F/A-18 Hornet fighter as well as a simpler cantilever beam design. Lim adopted the framework of stochastic programming with recourse in order to optimize the initial design of a system while probabilistically considering events that could unfold one period in the future. Lim suggested a combination of deterministic scenario-based optimization, stochastic programming, and interactive decision support tools to design evolvable systems using a 9-step process named EvoLVE.

The work of Christian and Olds[24-25] is another recent example of aerospace literature considering flexibility. In their work, Christian and Olds describe flexibility in terms of a system's ability to move between different end states in a lawful state space. An example application evaluates two competing human exploration architectures in terms of their ability to easily achieve extended lunar missions. Three state variables describe the performance requirements of the extended lunar mission,[†] and a Difficulty Scale for Evolvability Analysis (DSEA) is formulated to permit expert judgement to rate the difficulty (on a 1-3-9-27-81 scale) of evolving each architecture to meet various second-period performance states. The authors observed that "a single metric cannot capture the sensitivity of an architecture's capability to evolve" since that capability depends on the final evolved state that is desired.

Finally, in 2006, Silver and de Weck[28-29] proposed an analysis of evolvability based on expansion of a network of system operating and switching costs through several time periods. A set of deterministic exogenous demand

[†] Contrary to Jones and Ostroy, whose state-space "positions" appear to refer to future options, the state space of Christian and Olds is defined by the *performance* of those future options.

scenarios was assumed, and an optimizer was used to find the least-cost path through the network for each scenario. The method was referred to as a time-expanded decision network (TDN) and was applied to selection of a NASA heavy-lift launch vehicle. One notable limitation to the method is its single-objective and deterministic approach: Since the exact present and future demands of each scenario are known in advance to the decision-maker (or optimizer), paths through the time domain are able to fully specify an optimal solution. No explicit consideration is given to the possibility that a decision-maker will make choices in part to hedge against uncertain future events.

D. Flexibility in Aerospace Engineering Practice: An Example

In May 2005, NASA Administrator Michael Griffin commissioned the Exploration Systems Architecture Study (ESAS)[37] to recommend an architecture to support sustained human and robotic lunar exploration. In its trade studies, ESAS used five categories of figures of merit, one of which was Extensibility/Flexibility. Within this category were considerations of lunar mission flexibility, Mars mission flexibility, extensibility to other exploration destinations, commercial extensibility, and national security extensibility. ESAS characterized these flexibility considerations in terms of qualitative high (green), medium (yellow), low (red) ratings based on expert judgement. One example of these qualitative ratings for an evolved expendable launch vehicle (EELV) derived crew launch vehicle (CLV) is shown in Fig. 2.

	LV	EELV-derived CLV			
		Atlas V HLV New Upper Stage Human-Rated	Atlas Evolved Crew	Atlas Phase 2 Crew	Delta IV HLV New Upper Stage Human-Rated
		2	5.1	9	4
FOMs	Probability of Loss of Crew	1 in 957	1 in 614	1 in 939	1 in 1,100
	Probability of Loss of Mission	1 in 149	1 in 79	1 in 134	1 in 172
	Lunar Mission Flex-ibility				
	Mars Mission Extensibility				
	Commercial Extensibiity				
	National Security Extensibility				
	Cost Risk				
	Schedule Risk				
	Political Risk				
	DDT&E Cost	1.18	2.36	1.73	1.03
	Faciities Cost	0.92	0.92	0.92	0.92

Figure 2. Sample summary of figure of merit ratings for concepts in the ESAS report.[37] *Note the qualitative red/yellow/green ratings for flexibility.*

Aside from the primarily academic literature surveyed above, the ESAS methodology largely reflects of the state of the practice in designing for system flexibility today. The approach has positive qualities in that it considers flexibility during conceptual design process, and it does so with the recognition that flexibility must be traded against other objectives such as cost. As a result, this approach is amenable to application of standard multi-attribute decision-making techniques. However, this approach has disadvantages in its subjectivity and its use of a Likert-like qualitative scale with no physical units. This inhibits the analysis' repeatability and allows substantial room for results to be disputed. More fundamentally, the method treats flexibility as a scalar metric of the same class as cost or performance; however, it might reasonably be argued that the decision-maker does not care about flexibility itself (in whatever units one chooses for it), but rather cares about the *effects* that designed-in flexibility may have on future cost or performance.

E. Gaps in the Literature

This section has surveyed a broad set of engineering and economics literature; in the process, certain gaps have become evident in current thinking on flexibility and in current methods to consider this property in system design:

- In much of the literature, there appears a tendency for engineers to consider flexibility as a system-dependent scalar quantity. This concept has driven the invention of numerous scalar measures for flexibility that are often subjective and expressed on a scale with no units or clear physical interpretation. Further, when or if these measures are used in trade studies, they imply that flexibility is a property of the system separate from all others (such as cost and performance measures). *However, the decision-maker likely has little interest in flexibility for the sake of flexibility: He or she cares about flexibility primarily because of cost and performance benefits it may enable in the future.*

- *Few existing methods for considering flexibility consider more than one period in the future.* While considering one future period is an important first step, it is only one period less myopic than the traditional single-period horizon. If a system or program is to be operated for many decades (as is often the case in the aerospace industry), the prudent decision-maker cares not only to consider options for the first time that requirements or environments change, but also for many subsequent changes.

- *Furthermore, of methods that do consider implications of flexibility more than one period into the future, few utilize stochastic models.* Some methods assume a deterministic schedule of future requirements, while others select a handful of deterministic scenarios upon which to evaluate the system of interest. However, the probability of any one scenario occurring may be nearly (or, if continuous random variables are involved, exactly) zero. Without an understanding of the underlying probabilities of transition between demand or requirement environments, it may be problematic to assume a handful of scenarios can properly represent the entire space of possible futures.
- While some existing methods (such as decision trees) permit valuation of the avenues of flexibility provided by a system, they typically operate by assuming a single expected-value objective function. *In reality, engineering design involves trades among multiple cost and performance metrics as well as measures of dispersion for these parameters when subject to a stochastically changing environment.*
- Finally, the flexibility literature contains little discussion about the policies that flexible system operators need use to decide whether to exercise the options provided by flexibility. Some appear to assume that the appropriate policy is to always modify the system to precisely meet the anticipated demand or requirement. However, this is a very special case, and it may be in the program's best interests not to meet this demand if it is likely to be transient, or to over-perform if doing so is likely to boost performance in a later period of high demand. *The policy by which the system will be operated is an important part of system design, especially for a flexible system.*

In summary, today there exists no quantitative, stochastic, multi-objective, and multi-period framework for integrating flexibility into space system design decisions. It is such a framework that this paper proposes, drawing from literature and tools from industrial engineering, aerospace engineering, and economics in order to operationally define flexibility and transform its consideration into a tractable problem of stochastic optimal control.

III. A Markovian State-Space Flexibility Framework

The gaps in the present literature listed in Section II.E suggest that at least four components are critical for a decision framework that integrates flexibility into space system design decision-making: First, a stochastic model for the evolution of system demand over multiple time periods must be developed; such a model must describe what a system may be expected to accomplish (or what a decision-maker may be rewarded for performing) in the future. Second, a set of candidate system designs or configurations must be developed that is valid for multiple time periods in the future; this describes the future options available to the decision-maker and is suggested by the two-period state-centric notion of flexibility in the literature. Quantitative performance measures are required to evaluate how well the configuration that is fielded at a given time fulfills the demand or mission requested of it; in some scenarios, multiple performance measures will be required to capture trades among multiple objectives. Finally, since decisions regarding which system(s) to develop and field next must be made at multiple future time periods, a process must exist for integrating flexibility considerations into these decisions in an easily interpretable manner. Since the framework developed in this paper is intended to be used by decision-makers facing an immediate system selection problem, of particular interest is identification of the best system(s) to develop and field initially. These components are illustrated graphically in Fig. 3.

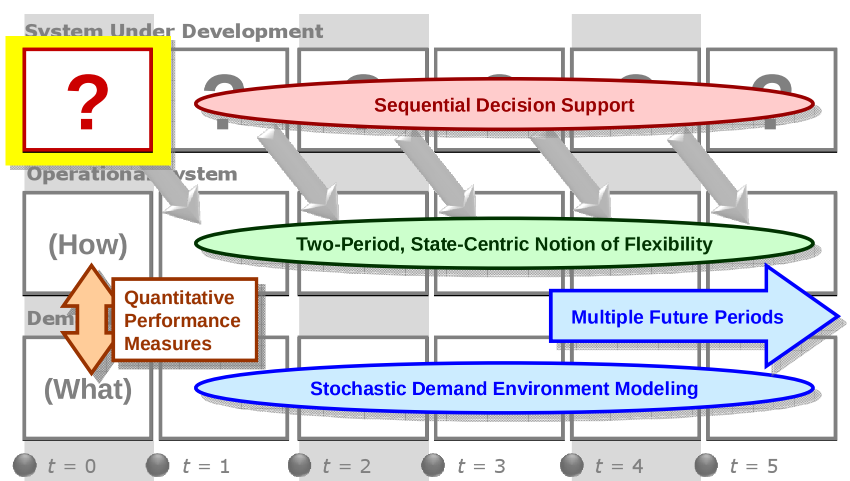

Figure 3. Critical components for decision frameworks addressing gaps in flexibility literature.

To accommodate these requirements, this paper presents a framework consisting of five basic steps, outlined in Fig. 4. First, system configuration options are identified and costs of switching from one configuration to another are compiled into a cost transition matrix. Second, probabilities that demand on the system will transition from one mission to another are compiled into a mission demand Markov chain. Third, one performance matrix for each design objective is populated to describe how well the identified system configurations perform in each of the identified mission demand environments. Fourth, possible future sequences of system configurations are simulated and sequences that are Pareto-optimal in terms of the decision-maker's objectives are identified. In a complementary approach, the system decision problem is formulated as a multi-objective variant of a Markov decision process, and Pareto-optimal decision policies are identified. Finally, the paths and policies from the latter step are synthesized into a set of data to inform initial system selection.

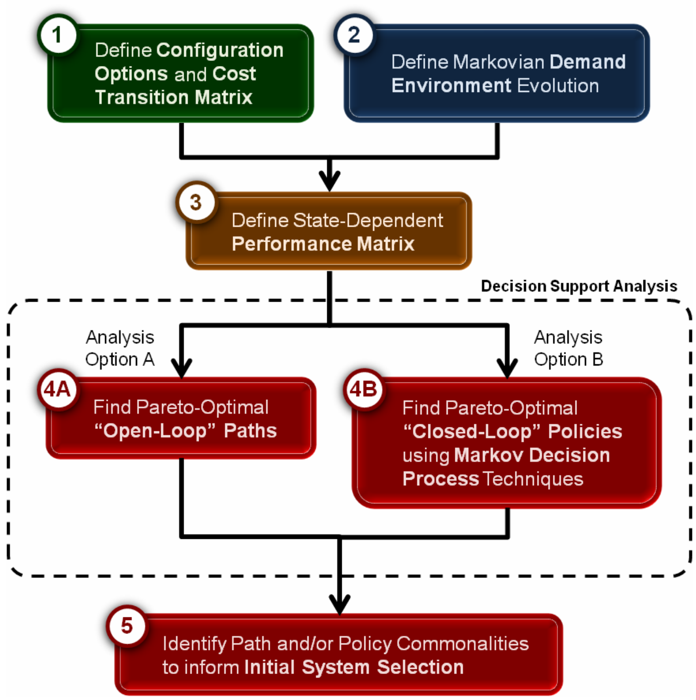

Figure 4. Five major steps of this paper's framework.

IV. Application: Design of a Distributed-Payload Satellite System

To illustrate this framework in a step-by-step manner, this paper poses a realistic example application in which design decisions must be made for a hypothetical multi-payload Department of Defense satellite system. Motivated by the distributed-payload monolith concept recognized by the DARPA F6 program (e.g., see Refs. 39-41), the following application illustrates how flexibility-related trades can be quantified for this basic fractionated spacecraft concept. Of particular interest is the answer to the following question: How can a systems engineer or analyst select the design of the satellite system initially such that it can optimally (or Pareto-optimally) respond to the uncertain future demands that may be placed upon it?

A. Step 1: Define Configuration Options and Cost Transition Matrix

As noted in Section II.B, in 1984 economists Jones and Ostroy[23] suggested, "Flexibility is a property of initial positions. It refers to the cost, or possibility, of moving to various second period positions." Similar views are supported elsewhere in the literature. Thus, step 1 of this proposed framework begins by defining: What are the possible "positions" of this satellite system?

1. Defining the Configuration Space

This paper's framework proposes that the "positions" of an engineering system are its possible configurations, or its possible design options. This choice for the position definition has the reasonable implication that given enough resources, the engineer or decision-maker can choose to field any particular system configuration (or be at any particular "position") in the future.

What it is that precisely defines these configurations is application-specific. In the case of this distributed-payload satellite system application, suppose that the decision-maker has the option of utilizing up to three specific payloads in any current or future system designs. One payload (PL1) provides detection of distress transmissions, another (PL2) provides high-bandwidth communications, and a third (PL3) provides high-resolution imagery. Assumptions for mass, power, and pointing requirements for these payloads are shown in Table 1.[‡] Considering that these three payloads can be distributed among up to three on-orbit modules and that not all three payloads need be

[‡] This list of payloads is limited to three for demonstration purposes only and can easily be increased if a decision-maker wishes to consider additional candidate payloads.

included in the system design (i.e., that omitting payloads is a valid consideration), there exist 15 distinct configuration options. These configurations are represented graphically in Fig. 5 and, as noted in previous work[39], can be decomposed into subsets of configurations described by Bell numbers. Starting from the bottom, configurations 11-15 represent all possible ways of distributing three payloads among between one and three modules (i.e., from monolithic to fully fractionated). Configurations 5-10 cover all possible ways of distributing combinations of two payloads among up to two modules. Configurations 2-4 are the single-payload satellite system options, and Configuration 1 indicates the option to field no system at all.

Even at this early point in the process, enumeration of the designs within the configuration space reveals two extremes in approaches for evolving the system to meet future needs: The most modular (but in the long term, potentially costly) approach would be to launch new single-payload modules as new payloads are needed. A robust (but in the short term, potentially wasteful) approach would be to launch a single spacecraft with all three payloads, betting that all capabilities will eventually be required. A number of approaches fall between these extremes, and an important goal is to find the best possible sequence of configurations over the system's time horizon, given the uncertainty in future demand or requirements. One of the most important results of this search is eventual identification of the best possible initial design (i.e., what the decision-maker should build at the start of the program).

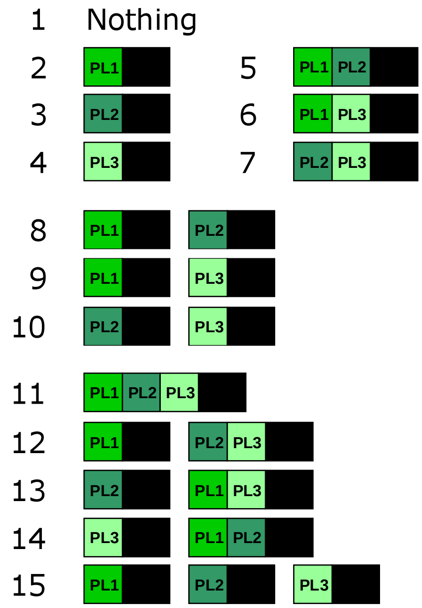

Figure 5. Possible system configurations. *Each distinct rectangular block represents a free-flying module. The payloads inside each module are indicated in green.*

Table 1. Assumed payload characteristics for example design.[42-46]

Payload No.	Payload Description	Flight Heritage	Mass (kg)	Power Requirement (W)	Pointing Requirement (deg.)
1	Search & Rescue Repeater	NOAA-N	24.0	53	1.00
2	LEO Transponders	Orbcomm	8.4	10	5.00
3	High Resolution Imager	NigeriaSat-2	41.0	55	0.01

2. Defining the Cost Transition Matrix

Recalling that flexibility "refers to the cost, or possibility, of moving to various second period positions" [23], to proceed it is necessary to incorporate cost information in addition to information on the composition of each system configuration. For space systems, these costs typically consist of development and operations costs. Here, operations costs will refer to the total costs required to operate the currently-fielded configuration over the coming time period. Development costs will refer to the total costs required to design, develop, produce, and launch the components needed to transition from the current configuration to a new configuration over the coming time period.

In this scenario, suppose the decision-maker encounters a decision point every 30 months. At these points, a decision must be made regarding which of the 15 system configurations to develop and then field 30 months later. Demand for payload services in each 30-month operations period is uncertain a priori and materializes after development, with the possibility that it will then change in subsequent period (see Fig. 6).

Thus, the decision-maker has control over the system configuration but not the demand environment at each time step. However, at each decision point, the control that the decision-maker chooses to exercise comes at a certain cost. For example, if the decision-maker is at the second decision point and has Config. 2 already on-orbit, in order to transition to Config. 8 he/she would need to expend the appropriate resources to develop and launch a new module. In addition, he/she must simultaneously pay for the operation of the current on-orbit system.

These transition costs can be represented in matrix form. First, a development (or nonrecurring) cost matrix C_{dev} accounts for the one-time costs required to develop and produce one system given that another system already exists. This cost, which can also be considered a switching cost, is the cost most central to the notion of flexibility and may be computed through application-specific cost estimating relationships. In this case, application of the GT-FAST fractionated architecture synthesis tool[39,47] using the payload assumptions of Table 1 for a 10-year design lifetime in a 410 km circular orbit produces the transition cost estimates in Table 2. These costs include appropriate spacecraft subsystem development and first-unit production, program management and systems engineering, software, ground segment development, launch, and assembly, test, and launch operations (ATLO).

Importantly, note that Table 2 accounts for the fact that free-flying modules for the next-period architecture need not be developed or produced if they exist already within the on-orbit cluster. The most obvious manifestation of this is that the diagonal of matrix C_{dev} consists entirely of zeros; this signifies the intuitive fact that it costs nothing to develop configuration *i* given that configuration *i* already exists. Similarly, note that no development costs are required to downgrade a configuration, such as a transition from Config. 15 (which, as shown in Fig. 5, includes three single-payload modules) to Config. 2 (which consists of only the PL1 single-payload module). This highlights a simplifying assumption within the data of this particular matrix that the cost to shut down or decommission a module is zero; however, given proper decommissioning cost models, this information could easily be included in C_{dev}.

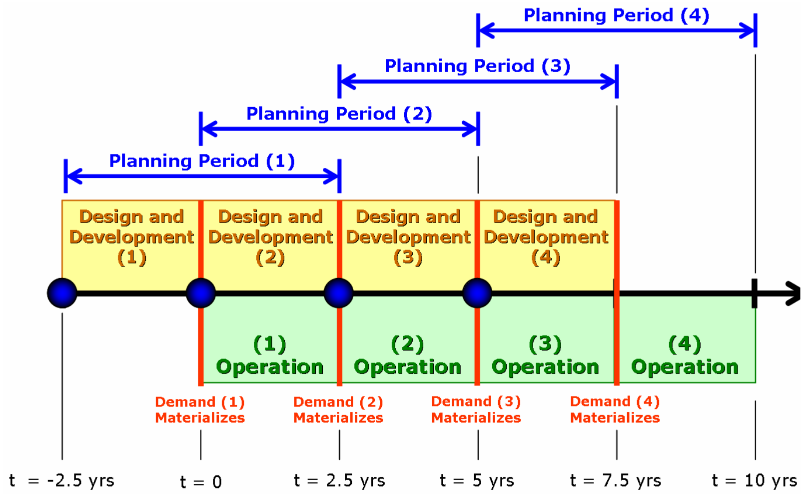

Figure 6. Planning periods and decision points (in blue) over a 10-year time horizon.

Table 2. Development cost transition matrix, C_{dev} (data in $FY08M).

From Configuration \ To Configuration	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	169	131	184	175	200	189	197	252	212	204	257	228	258	280
2	0	0	36	89	80	105	94	36	89	117	109	94	134	163	117
3	0	75	0	89	80	105	94	75	158	89	109	163	105	163	158
4	0	75	36	0	80	105	94	103	75	36	109	163	134	80	103
5	0	75	36	89	0	105	94	103	158	117	109	163	134	89	186
6	0	75	36	89	80	0	94	103	158	117	109	163	36	163	186
7	0	75	36	89	80	105	0	103	158	117	109	75	134	163	186
8	0	0	0	89	80	105	94	0	89	89	109	94	105	163	89
9	0	0	36	0	80	105	94	36	0	36	109	94	134	80	36
10	0	75	0	0	80	105	94	75	75	0	109	163	105	80	75
11	0	75	36	89	80	105	94	103	158	117	0	163	134	163	186
12	0	0	36	89	80	105	0	36	89	117	109	0	134	163	117
13	0	75	0	89	80	0	94	75	158	89	109	163	0	163	158
14	0	75	36	0	0	105	94	103	75	36	109	163	134	0	103
15	0	0	0	0	80	105	94	0	0	0	109	94	105	80	0

Second, a recurring cost matrix C_{rec} shown in Table 3 accounts for operations and any production beyond the first unit.[§] In this example application, first-unit production costs are the only applicable production costs, so the costs within this matrix are functions only of the row, i.e., the configuration that is operational over the length of the

[§] In some instances, the analyst may wish to account for all of production within the recurring cost matrix, since even one-time production for a unique flight unit is traditionally bookkept as a recurring cost. In the present application, one-time module production costs are considered more closely related to the one-time development costs and are accounted for in the development cost matrix.

coming 30-month time period. These costs are also estimated using the GT-FAST tool, which draws upon a publicly-available NASA mission operations cost model.[48]

Table 3. Recurring cost transition matrix, C_{rec} (data in $FY08M).

	To Configuration														
From Configuration	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
3	16	16	16	16	16	16	16	16	16	16	16	16	16	16	16
4	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
5	21	21	21	21	21	21	21	21	21	21	21	21	21	21	21
6	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23
7	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22
8	23	23	23	23	23	23	23	23	23	23	23	23	23	23	23
9	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29
10	25	25	25	25	25	25	25	25	25	25	25	25	25	25	25
11	24	24	24	24	24	24	24	24	24	24	24	24	24	24	24
12	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29
13	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26
14	29	29	29	29	29	29	29	29	29	29	29	29	29	29	29
15	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31

Summing C_{dev} and C_{rec} from Tables 2 and 3 yields the total cost transition matrix C in Table 4. Each element $c_{i,j}$ of this matrix specifies the total cost incurred over a subsequent 30-month time period as the result of the decision to transition from developing configuration i to developing configuration j. For example, to transition from Config. 2 to Config. 8 requires developing, producing, and launching the module containing PL2 as well as operating the current Config. 2, for a total transition cost $c_{1,7}$ = $56 million.

Table 4. Total cost transition matrix, C (data in $FY08M).

	To Configuration														
From Configuration	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	169	131	184	175	200	189	197	252	212	204	257	228	258	280
2	20	20	56	110	101	126	115	56	110	138	130	115	154	184	138
3	16	91	16	106	97	122	111	91	174	106	125	179	122	180	174
4	22	96	58	22	102	127	116	124	96	58	131	184	155	102	124
5	21	96	57	110	21	126	115	123	178	138	130	184	155	110	207
6	23	98	59	113	104	23	118	126	181	141	133	186	59	187	209
7	22	97	58	112	103	128	22	125	180	140	131	97	156	186	208
8	23	23	23	113	104	129	118	23	113	113	132	118	129	187	113
9	29	29	65	29	109	134	123	65	29	65	138	123	162	109	65
10	25	99	25	25	105	130	119	99	99	25	134	187	130	105	99
11	24	98	60	113	104	129	118	126	181	141	24	186	157	187	209
12	29	29	65	118	109	135	29	65	118	147	138	29	163	193	147
13	26	101	26	116	107	26	121	101	184	116	135	189	26	190	184
14	29	104	65	29	29	135	124	132	104	65	138	192	163	29	132
15	31	31	31	31	112	137	126	31	31	31	141	126	137	112	31

3. *Analyzing the Cost Transition Matrices*

The data represented by the cost transition matrices can be analyzed, visualized, and related to flexibility in several useful ways. First, the relative trade between system initial costs and the switching costs (or one-time development costs) of Table 2 can be visualized as in Fig. 7. In this figure, each vertical line indicates the range of switching costs from a given configuration, defined by the rows of Table 2. Solid dots indicate minimum and maximum values, and triangles indicate median values. Each vertical line is located horizontally at the cost needed to develop the configuration from scratch (in this case, Config. 1). For example, if no system currently exists and a decision-maker chooses to develop Config. 5 (involving a single module with PL1 and PL2 on board), a cost of $175 million is incurred (on the *x*-axis), and the cost to switch configurations in the future varies from $0 to $186 million, depending on which future configuration is chosen. In contrast, if the decision-maker instead chooses to develop Config. 15 (involving three payloads among three modules), a cost of $280 million is initially incurred, and the cost to switch configurations in the future varies from $0 to $109 million. Thus, to some extent Fig. 7 empirically confirms the intuitive trend that future switching costs can often be reduced by earlier investments.

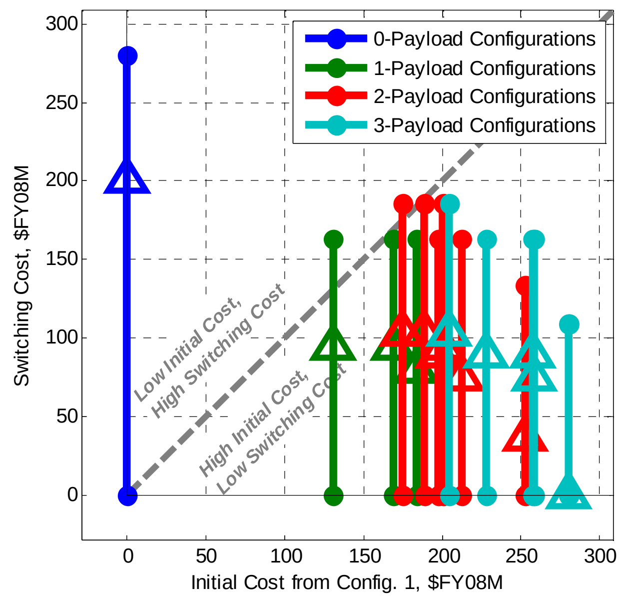

Figure 7. Switching cost vs. initial cost from Config. 1. *Vertical lines indicate ranges of switching costs from each configuration; some overlap. Solid dots indicate minima and maxima, and triangles indicate median values.*

Second, the data from the total cost transition matrix (Table 4) can be visualized directly in the context of the two-period state-centric notion of flexibility mentioned earlier. For this visualization see Fig. 8. Here, each node in each of the three plots represents one of the configurations considered in the design space. Each node is named S_X, where X is the configuration number from Fig. 5, and has a color indicative of the number of on-board payloads (consistent with the colors of Fig. 7). Above each of the three plots is a budget, and for every element of the total cost transition matrix less than or equal to the given budget, a directed link is drawn. In cases where the total cost on the diagonal of the matrix is less than or equal to the budget, a dark circle is drawn around the appropriate node. For example, the middle plot of Fig. 8 shows that, if the currently-fielded architecture is Config. 12, a $50 million budget for a given 30-month period would allow the decision-maker to transition to Configs. 1, 2, or 7, or to remain in Config. 12. In cases where no links or dark circles are associated with a configuration, the available budget is insufficient even to support operation of the current configuration into the next period.

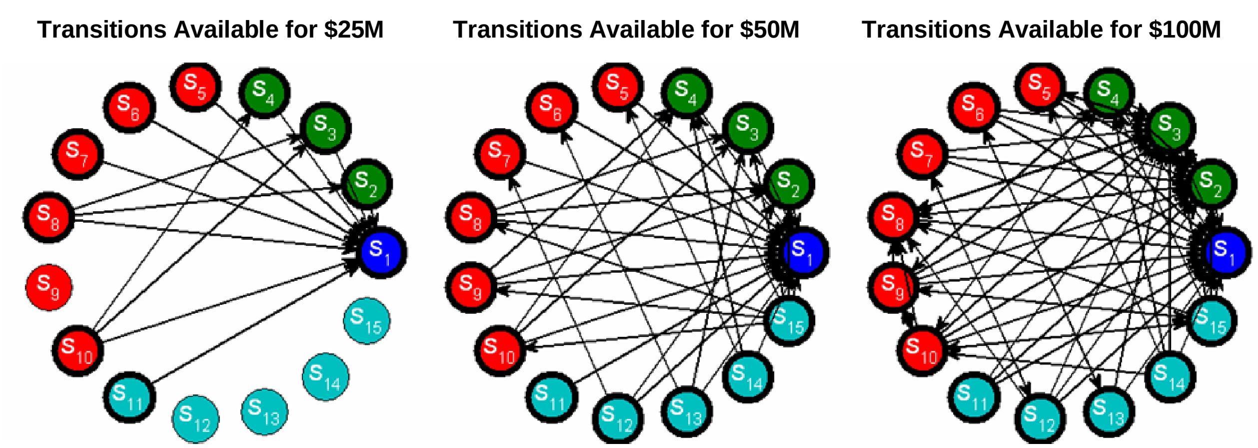

Figure 8. Available configuration transitions for three example 30-month budgets.
Self-transitions are available if a dark ring circles a given configuration. Colors indicate each

A natural observation from Fig. 8 is that, as budget is increased, more links become available. That is, as the decision-maker has more resources available, more options exist. The total number of links in the graphs of Fig. 8 increases from 23 at the $20 million budget to 47 at the $50 million budget and 78 at the $100 million budget. Eventually, at a large enough budget, all 225 links would appear. Linking this to the two-period state-centric concept of flexibility, a clear indicator of the flexibility of a given configuration *i* is the number of links or transitions available to it for a given budget *b* (the number of "outs" available, which will be denoted $\Phi_i(b)$).

This indicator is plotted in Fig. 9. The figure shows the number of available transitions as a function of available budget, where data for each configuration is represented by a single line. For example, the figure illustrates that for a per-period budget of $50 million, Config. 1 (the "nothing" configuration) has Φ = 1 transition available, Configs. 2-7 and 11 each have Φ = 2 available transitions, Configs. 8-10 and 12-14 each have Φ = 4 available transitions, and Config. 15 has Φ = 8 available transitions. It also shows that by a budget of $300 million, any configuration can be reached from any other configuration since all configurations have 15 available transitions.

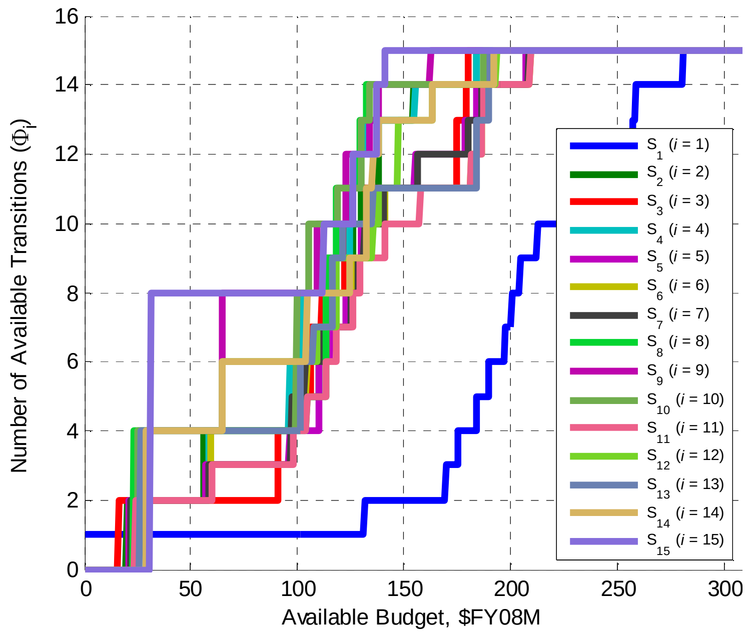

Figure 9. Available configuration transitions as a function of the available 30-month budget.

Figure 9 highlights a few interesting transition characteristics for the configurations in the design space defined in Fig. 5. If the quantity Φ is interpreted as a surrogate measure of flexibility, then it is easily seen that Config. 1 is significantly less flexible than any other configuration over most of the budget range plotted in Fig. 9. For Config. 1, the first available transition to another configuration occurs at $131 million; for the same budget, other configurations can already make between 9 and 13 transitions. This occurs because Config. 1 has no capabilities that it can leverage to easily transition to other configurations, and all capabilities must be developed from scratch. It is also relevant to note that the three-payload monolith, Config. 11, which has no modules in common with other configurations, tends to have fewer transitions available than most other configurations at most budget levels. On the other hand, Config. 15 (the fully fractionated design) very quickly attains a large number of available transitions as budget increases; this configuration is the first to reach 8 transitions and the first to attain the ability to make all 15 available transitions. This occurs because Config. 15 consists of three single-payload modules that can easily be used as pieces of other configurations; from Config. 15, the only modules that must be developed to reach other configurations are the two- or three-payload modules.

In terms of number of transitions, the other configurations within the design space generally fall between the bounds of Configs. 1 and 15. All illustrate that Φ is a monotonically increasing function of budget, which implies that any given configuration's flexibility increases with available budget. However, examples also can be found to illustrate that the *relative* flexibility between configurations is also a function of available budget. For example, at a budget of $25 million, Config. 8 has four available transitions while Config. 15 has none. In other words, at a budget of $25 million, it is reasonable to make the statement that Config. 8 is more flexible than Config. 15. However, at a budget of $50 million, Config. 8 still has four available transitions while Config. 15 can make eight transitions. At this budget level, Config. 15 is more flexible than Config. 8, and the relative flexibility of these configurations has reversed. The reason for this "flexibility reversal" becomes evident when it is recalled that the cost transition matrix accounts for both development and recurring operations costs: When budget resources are scarce, operating a high-capability configuration (like Config. 15) consumes funds that would otherwise be available for developing the components needed to transition to another configuration. However, as financial resources become more abundant, more capable configurations become more flexible because they already possess capabilities transferrable toward the development of other configurations.

4. Key Observations from Step 1

This step of the framework has shown that the two-period state-centric notion of flexibility can be adapted to apply to configuration changes for space systems, with particular emphasis on a distributed-payload satellite. A cost transition matrix was formed and used to visualize the options that exist for changing the system as a function of available budget. If a single, relatively constant per-period budget is likely to exist for the foreseeable future, that budget can be selected and a diagram such as one of the graphs in Fig. 8 can be useful in tracing possible configuration pathways. If the available budget is likely to be subject to change or partially under the control of the decision-maker, the available transitions can be plotted as a function of budget to determine if additional budget would make a substantial difference in the available options. Analysis of these graphs and associated data illustrate how budget itself can drive whether one configuration is more flexible than another.

At the conclusion of this step, it is reasonable to ask: From this information, what conclusions can be drawn about the best initial system configuration to select? Unfortunately, none. To do so requires overcoming two limitations of considering only configurations and cost over a two-period time interval. First, the time horizon of the analysis must be expanded to more than two periods to avoid potentially myopic decision-making. Second, the benefits of being in a given configuration at a given time must be quantified. A limitation of considering only the number of available transitions metric is that it contains no information about the value of each configuration in each future time period. As a result, it is possible to manipulate this metric to make certain configurations appear relatively more or less desirable by either (1) including in the state space a large number of physically similar configurations or (2) including in the state space a large number of configurations that are unlikely to have any value in the future. These limitations are resolved in the following two steps of the framework.

B. Step 2: Define Markovian Demand Environment Evolution

Typically, the performance of a system depends upon the environment in which it operates. Thus, while Step 1 of this framework focuses on defining the system itself and its available configuration states, the environment in which the system operates has not yet been discussed. Step 2 fills this gap by proposing a model for the evolution of the environment. Unlike the configuration state, which is under the control of the decision-maker, the environment state will characterize the demands placed on the system at any given time, which inherently is not under the control of the decision-maker and evolves stochastically.

As mentioned in Section IV.A.1, up to three specific payloads are available for any current or future designs of the distributed-payload satellite system. One payload (PL1) provides detection of distress transmissions, another (PL2) provides high-bandwidth communications, and a third (PL3) provides high-resolution imagery. In terms of defining the demand environment, it is reasonable to expect that future demand may exist for the satellite system to provide any combination of these three services. For example, in one time period, only high-bandwidth communications may be required, and in another, both high-resolution imagery and high-bandwidth communications may be needed. Thus, there exist eight distinct demand environment states, indicated by the axes in Table 5. Note that these environment states are mutually exclusive and, for example, “1” should be interpreted as “1 only” and “1+2” should be interpreted as “1+2 only”.

It is also reasonable to expect that the evolution of demand for these services through time is unlikely to be properly modeled by a time series of independent random demand environments. Rather, a subsequent period’s demand likely depends in part upon the current demand, a dependence that can be captured using a Markov chain stochastic model. Formally, a Markov chain is an ordered set of discrete random variables (e.g., $\{Q(t)\}$) for which the probability that $Q(t)$ takes some value σ depends only on the value of $Q(t\text{-}\Delta t)$, i.e., Q in the previous time period. Thus, in a Markovian stochastic process, the past influences the future only through the

Table 5. Assumed demand environment transition probability matrix. *Note that, in the demand environment naming convention, 1 indicates demand for distress transmission detection , 2 indicates demand for high-bandwidth communications, and 3 indicates demand for high-resolution imagery services.*

From Demand Environment \ To Demand Environment	None	1	2	3	1+2	1+3	2+3	1+2+3
None	0.30	0.05	0.13	0.30	0.02	0.05	0.13	0.02
1	0.20	0.15	0.09	0.20	0.06	0.15	0.09	0.06
2	0.10	0.02	0.23	0.15	0.05	0.03	0.35	0.07
3	0.10	0.08	0.07	0.23	0.05	0.19	0.16	0.12
1+2	0.05	0.07	0.20	0.03	0.28	0.05	0.13	0.19
1+3	0.05	0.05	0.05	0.20	0.05	0.20	0.20	0.20
2+3	0.05	0.04	0.12	0.12	0.09	0.09	0.27	0.22
1+2+3	0.02	0.02	0.08	0.08	0.08	0.08	0.32	0.32

present state.[**] The conditional probabilities P[$Q(t) = \sigma \mid Q(t- \Delta t) = \varsigma$] with which values of Q at time t-Δt evolve to other values of Q at time t are organized in a probability transition matrix.

The particular probability transition matrix assumed for this example is shown in Table 5. Ideally, this matrix would be populated using a set of expert judgements regarding future demand behavior or, if they exist, probabilities based on historical data. In this notional example, the author's judgement was used to select values that reflected a high likelihood that a current demand would be maintained (e.g., if high-resolution imagery is demanded in the current period, it would be likely to also be demanded in the next period) and tended to place lower probabilities on the need for dedicated distress transmission detection services. The probabilities in Table 5 also reflect an assumed conditional independence in the evolution of demand for each individual service; for example, given a particular demand $Q(t$-$\Delta t)$, the probability of demand evolving for all three services (1+2+3) is equivalent to the product of three underlying probabilities that are conditional on $Q(t$-$\Delta t)$ but reflect the likelihood that demand evolves to each of the services individually. It is important to emphasize, however, that the particular probabilities in Table 5 are illustrative and can easily be substituted with if more data or other expert judgements become available.

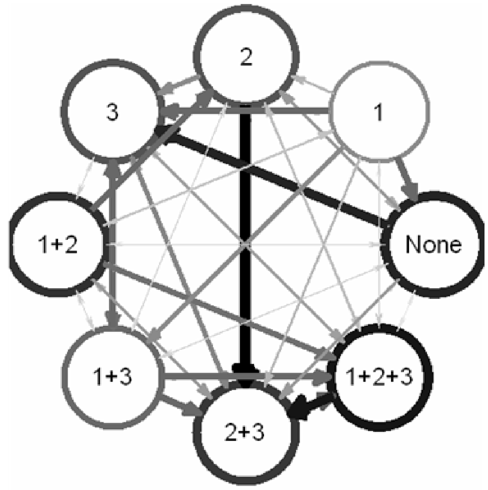

Figure 10. Visualization of the demand environment Markov chain described by Table 5.

The Markov chain of Table 5 can be visualized as a set of demand environment states as in Fig. 10. In this figure, high-probability transitions are represented as thick dark links and low-probability transitions are represented as thin light links. The likelihood of self-transitions (along the diagonal in Table 5) are indicated by the darkness and thickness of rings around each state. Thus, for example, this figure immediately allows identification of the highest-probability and lowest-probability transitions in the Markov chain and demand environment evolution.

C. Step 3: Define State-Dependent Performance Matrix

Linking the on-orbit configuration to the demand environment is a matrix that specifies the amount of reward (e.g., revenue or an accumulated performance measure) earned in each time period as a function of the demand environment and system configuration in that period. The application here uses the matrix in Table 6, which specifies the number of demanded services that are performed given a particular configuration operating in a particular demand environment. For example, if the demand in one time period is for imagery and communications (column 7) and the vehicle on-orbit is in Config. 15 (the 3-payload fully-fractionated option, row 15), the decision-maker accumulates the performance of two demanded services. As a result, the decision-maker is incentivized to place payloads in orbit that will meet demand for services.

Table 6. Performance matrix quantifying the number of demanded services performed in a given time period.

		Demand Environment State							
		None	1	2	3	1+2	1+3	2+3	1+2+3
Configuration State	1	0	0	0	0	0	0	0	0
	2	0	1	0	0	1	1	0	1
	3	0	0	1	0	1	0	1	1
	4	0	0	0	1	0	1	1	1
	5	0	1	1	0	2	1	1	2
	6	0	1	0	1	1	2	1	2
	7	0	0	1	1	1	1	2	2
	8	0	1	1	0	2	1	1	2
	9	0	1	0	1	1	2	1	2
	10	0	0	1	1	1	1	2	2
	11	0	1	1	1	2	2	2	3
	12	0	1	1	1	2	2	2	3
	13	0	1	1	1	2	2	2	3
	14	0	1	1	1	2	2	2	3
	15	0	1	1	1	2	2	2	3

[**] If it is necessary to build additional memory into the process, it is possible to do so by expanding the chain's state space (i.e., the definition of the possible values of Y).

It is also worth noting that, although the present application adopts just one performance metric (and thus one performance matrix), multiple such matrices can be defined for any cumulative performance metrics of interest to the decision-maker. For example, a decision-maker may also be interested in a cumulative binary metric that indicates a 1 or 0 in each time period depending on whether performance demands were fully met; over the long term, such a metric would indicate the percentage of time that the system fully meets the demands placed upon it.

D. Step 4: Decision Support Analysis

With configuration transitions, demand environment transitions, and a performance matrix defined, there now exists enough information to run a simulation and begin to answer the question of what is the "best" initial configuration the decision-maker can choose. Using Fig. 6 as a framework for a simulation timeline, one time period before a configuration is fielded (in this distributed-payload satellite example, at t = -2.5 years), a decision-maker must choose which system configuration to initially design, develop, and produce. At t = 0, the system that had been developed over the previous time period is fielded, and a demand environment materializes. At this point, the system operator must make use of the currently operational system in attempting to fulfill the current demand. Meanwhile, the decision-maker must choose which configuration to design, develop, and produce over the coming period. The cycle then repeats for as many periods as fills the time horizon under consideration. In this case, the time horizon of interest is 10 years of operation.

The decision support analysis in this step is divided into two complementary analysis options. The first option, in which Pareto-optimal paths are identified, is simpler to implement and conceptually similar to long-term scheduling and roadmapping analysis. The second option, in which Pareto-optimal policies are identified, is a more complete consideration of the problem and is akin to developing an optimal "playbook" of what actions to take given all possible future evolutions of the environment.

1. Step 4A: Find Pareto-Optimal "Open-Loop" Paths

One question that Fig. 6 prompts is: What configuration should the decision-maker choose to develop at each time increment? In other words, what configuration should be selected for each of the yellow design and development blocks in Fig. 6? The answer is not obvious, especially since the demand environment evolves stochastically. For example, the decision-maker who wishes to be able to fulfill whatever demand the next period may bring would choose to build the most capable system possible, but this would come at substantial initial expense. The decision-maker who would gamble that tomorrow's demand will be the same as today's would develop few or no new architectural components and in doing so save significant resources; however, this would come with the inability to perform if the next period's demand materializes to require greater capability. Furthermore, whether one period's decision is best (e.g., high-reward or low-cost in the long run) is likely to be dependent on other decisions throughout the system lifetime. In the flexibility problem, it is in general necessary to consider all future decisions within a given time horizon in order to judge the appropriateness of any single decision. While this presents a unique difficulty within the realm of space system conceptual design, once complete it presents an automatic solution to the question of which configuration to select initially: The appropriate configuration to select initially is the first configuration decision from the "best" time-ordered sequence of decisions.

In this example, posing the problem such that we wish to find the optimal sequence of the four development decisions (each decision of which implies a selection among the 15 configuration options) means that there exist 15^4 = 50,625 possible sequences (or paths). Since the configurations on these paths are identified by the time on the clock at which they are chosen, this type of specification will be referred to as an open-loop path.

Assuming an initial condition at t = -2.5 years in which the operational configuration is nothing (Config. 1) and there is demand for none of the services (the "None" environment), one approach to solving this problem is to simulate all 50,625 paths subject to the stochastically-changing demand environment and identify which produces the "best" combination of performance and cost. Thus, for each of the paths, 1000 Monte Carlo simulations are run. At each time step of a simulation, the following events and computations occur:

1. Mission demand evolves stochastically according to the Markov chain estimate of Table 5.
2. The operator of the currently operational configuration attempts to use this system to fulfill the new mission demand, earning credit according to the performance matrix.
3. The decision-maker chooses which configuration to develop in the current time period and field in the next time period, incurring a cost according to the cost transition matrix. An available choice in any time period is to retain the current configuration, which requires no additional development resources.

A sample set of Monte Carlo simulation results is shown in Fig. 11. This figure shows the result of adopting a path representing an incremental buildup of capability in which Config. 4 (the PL3-only configuration) is fielded initially. In the next time period, a new module containing PL1 is launched, and PL2 is added in the third time period. The cluster of three modules operates until the end of the 10-year time horizon. Due to the simulation setup, a configuration decision must still be made in the final operational time period; since the cost of developing this final configuration will be incurred but no reward will be earned, Config. 1 (the "Nothing" configuration) is selected. As the bottom left portion of Fig. 11 shows, this particular path (denoted as [4 9 15 15 1], by the configuration decisions made at each step) is subject to a stochastically changing demand environment. The size of each yellow dot indicates the likelihood of demand being in a particular state (on the *y*-axis) at a given time (on the *x*-axis); note that all simulations begin in the "None" demand environment at t = -2.5 years, as specified by the initial condition. The right-hand portion of Fig. 11 indicates how per-period cost and performance vary over time. Note that the per-period cost decreases from $184 million for the initial investment to $31 million in the final operations period, and number of demanded services performed per period increases from zero to a mean of 1.67 in the final period. The total expected cost for this path over the time horizon is $407 million[††], and the total expected number of demanded services performed is 4.69.

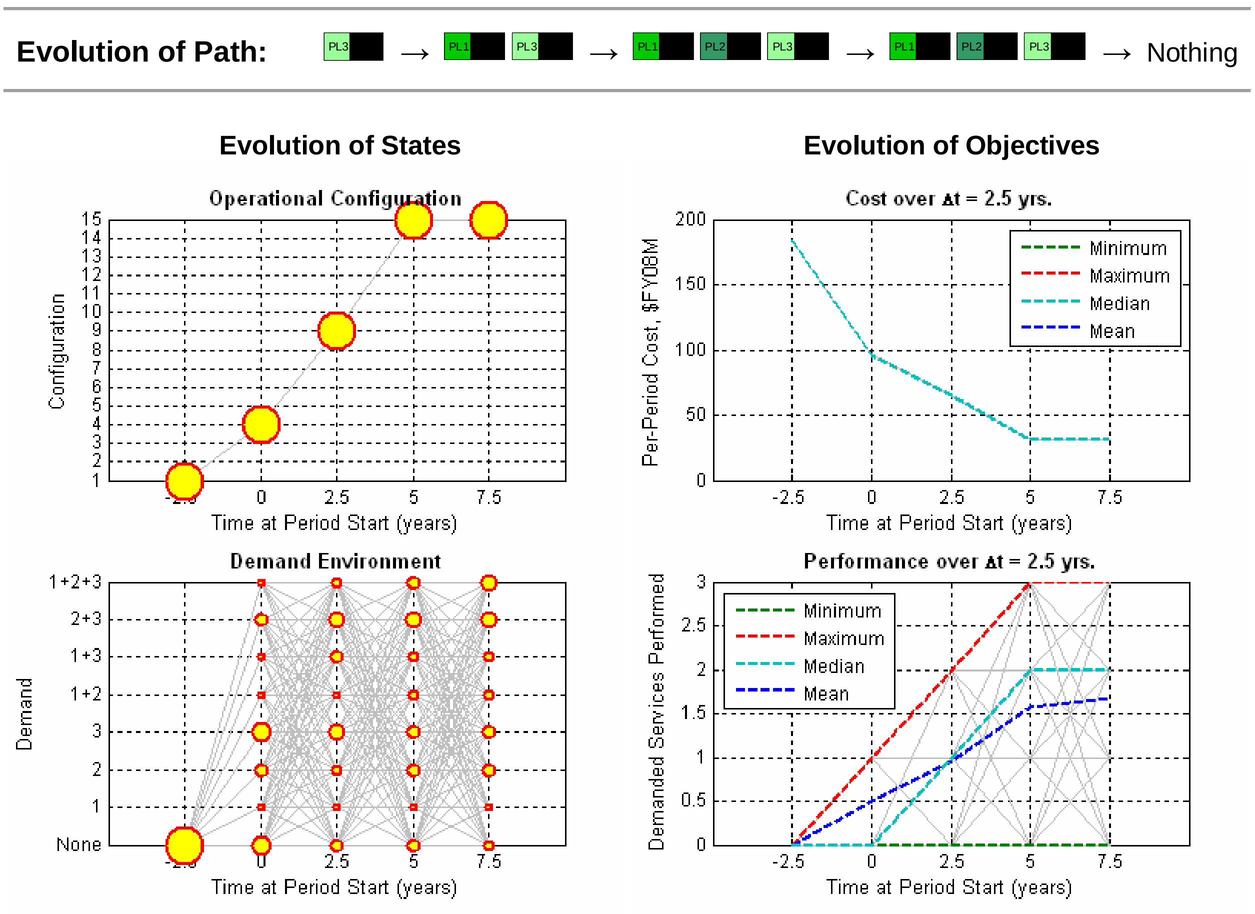

Figure 11. Evolution of configuration path [4 9 15 15 1], representative of an incremental capability buildup. *In the plots on the left, the size of circles indicate the relative number of Monte Carlo simulation cases that exist in a given configuration or demand environment state (on the y-axes) at a given time (on the x-axes). The plots on the right indicate the associated evolution of per-period cost and performance. In all plots, gray lines indicate transitions made in at least one simulation. Note configuration and cost are deterministic, since a path is specified.*

[††] Note that once a path is chosen, cost is fixed. As a result, the expected cost is equivalent to the minimum, maximum, and median costs across all path-based Monte Carlo simulations.

Obtaining results like those in Fig. 11 for each of the 50,625 possible paths allows the total expected performance to be computed and plotted against total cost for each path as in Fig. 12. In this figure, each blue "x" represents the total cost and performance of one path[‡‡]. Notice that, for the population as a whole, there is a general trend that, as more funds are invested, higher performance is expected. However, it is important to recall that the decision-maker has a choice of which path to select. As a result, if he or she cares primarily about total cost and expected total demanded services performed, it would make little sense to select a high-cost, low-performance point toward the lower right of the cluster. Rather, the decision-maker would prefer to choose among the set of nondominated points that comprise the Pareto frontier. This Pareto frontier, shown in red in Fig. 12, is composed of the set of possible configuration sequences for which one objective cannot be improved without the sacrifice of another.[§§] In this application, the frontier is comprised of just 12 of the 50,625 possible paths and helps to narrow the options considerably.

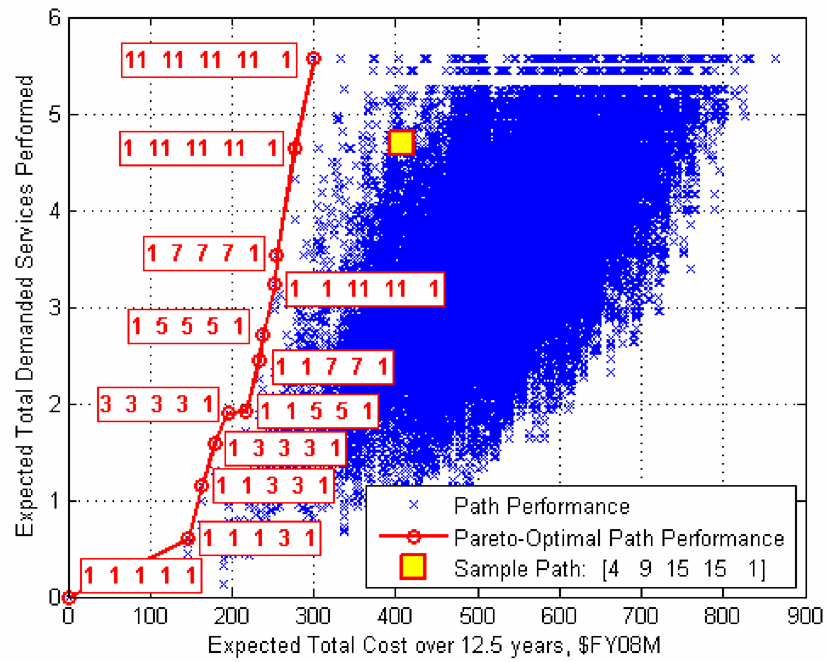

Figure 12. Trade between total demanded services performed and total cost for all open-loop paths. *Pareto-optimal paths are identified by 5-period configuration sequences listed next to red circles.*

Listed next to each of the Pareto-optimal points in Fig. 12 is its associated configuration path. Note that at the bottom left of the figure is the "do nothing" option in which Config. 1 is fielded for all time periods; this is cost-optimal but also provides the lowest possible performance. At the other extreme is the Pareto-optimal highest-performance option of fielding Config. 11, the three-payload monolithic satellite, for all time periods. The Pareto-optimal solutions between these two extremes involve developing Configs. 3, 5, 7, or 11, either immediately or after a 1-2 period delay. Notably absent from the frontier are the higher-cost multiple-module configurations.

One use of the data in Fig. 12 becomes evident when the sample path from Fig. 11 is overlaid as the yellow square in Fig. 12. Here it can be seen that the incremental path [4 9 15 15 1] is dominated by solutions on the Pareto frontier. In fact, one particular path, [1 11 11 11 1], accumulates near-identical performance for a total cost about $131 million (32%) lower. In this Pareto-optimal path, detailed in Fig. 13, the three-payload monolithic satellite is fielded after a one-period wait, during which time demand evolves toward an environment in which multiple services are demanded. Unlike the incremental path in Fig. 11, which exhibits a gradual decrease in per-period cost, the Pareto-optimal path in Fig. 13 exhibits an initial $204 million spike followed by $24 million in operations costs for three periods. As a result, this cost profile results in significant savings, and the system still performs well since all three payloads are available to fulfill all requested services at times in the future in which the environment has evolved to one in which multiple services tend to be demanded.

2. Step 4B: Find Pareto-Optimal "Closed-Loop" Policies

While straightforward and conceptually similar to an optimization of typical long-term scheduling and roadmapping efforts, the analysis presented in Step 4A has two principal disadvantages. First, for applications with large numbers of configurations and long time horizons, it may not be practical to enumerate all possible paths. For

‡‡ These totals are taken over the t = -2.5 year period (at which there is zero performance due to the initial condition) and the four subsequent periods.

§§ For further familiarization with Pareto optimality, Refs. 49 and 50 are recommended.

example, if the number of time periods in the present application were doubled, the number of possible paths would increase from 50,625 to over 2.6 billion and take several years of run time on a standard desktop computer. Second, assuming a set path for the entirety of the system's lifetime neglects the ability of the decision-maker to make choices mid-program in response to the evolution of the demand environment.

To overcome these limitations, Step 4B presents a complementary analysis that draws on a set of techniques from outside of the aerospace community that addresses this problem and is particularly well-suited for the state-space framework set forth in Steps 1-3. These techniques are associated with the class of stochastic control processes known as Markov decision processes (MDPs).

To define any MDP, it is necessary to first define (1) a set of states (or state space) S that describes the system of interest, (2) a set of decisions or actions A available from each state s, (3) transition probabilities $p(j|s,a)$ given that a particular decision a is made while the system is in state s, and (4) expected per-period rewards $h(s,a)$ associated with actions and/or states. In the case of MDPs on a finite time horizon, solutions typically exploit the computational efficiency of probabilistic dynamic programming, in which the overall maximization of a cumulative expected-value objective at the system's initial state and time $J_{s0,t0}$ is decomposed into a series of state-by-state, period-by-period maximization problems as specified in Eq. (4).*** The optimization is implemented starting from the final time period of interest and then working backward to the initial time t_0.

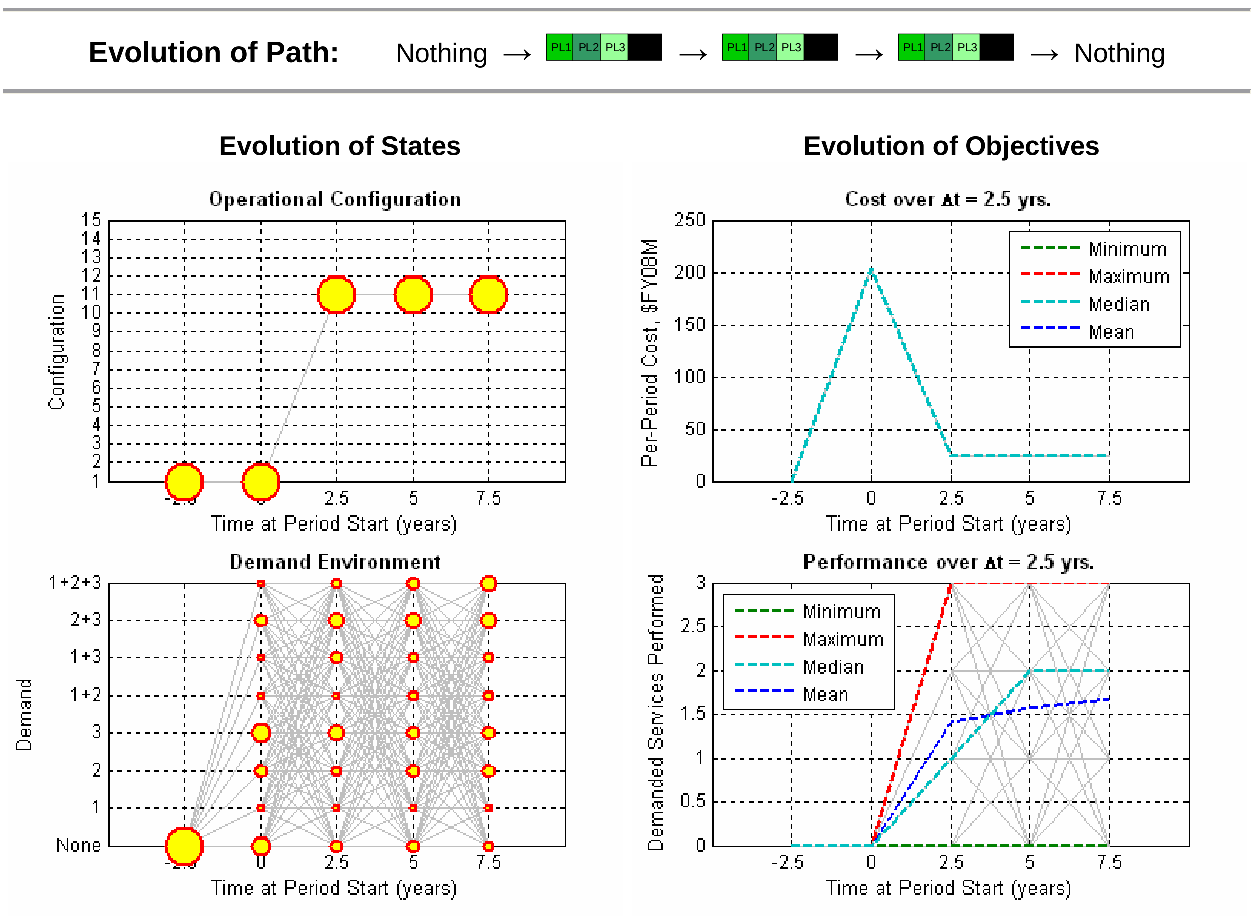

Figure 13. Evolution of configuration path [1 11 11 11 1], a Pareto-optimal path.
In the plots on the left, the size of circles indicate the relative number of Monte Carlo simulation cases that exist in a given configuration or demand environment state (on the y-axes) at a given time (on the x-axes). The plots on the right indicate the associated evolution of per-period cost and performance. In all plots, gray lines indicate transitions made in at least one simulation. Note configuration and cost are deterministic, since a path is specified.

*** The term β is a discounting factor. In this paper's analysis, β is set to unity.

$$J_{s,t} = \max_{a \in A(s)} \left(h(s,a) + \beta \sum_{j=1}^{S} p(j \mid s,a) J_{j,t+\Delta t} \right) \tag{4}$$

It may be evident that the state-space flexibility framework for system configurations established in Step 1, Markovian demand environment definitions of Step 2, and performance and cost information from both Steps 1 and 3 pose a problem that consists of states, possible configuration decisions, demand transition probabilities, and costs and performance rewards associated with transitions and states – all of which are the components of an MDP. However, two slight adjustments must be made to frame the present problem properly for an MDP:

First, the framework has so far used two *separate* state spaces. Step 1 introduced the configuration state space, and Step 2 introduced demand environment state space. To utilize an MDP formulation, the problem must be represented in a single state space. It is proposed that a total state be defined as the combination of the configuration and demand states (Total State = {Configuration State, Demand State}). In the distributed-payload satellite example, there are 15 configuration states × 8 environments = 120 total states, which Fig. 14 illustrates graphically. In this three-dimensional "spindle" of total states, each vertical layer represents a particular demand environment and each column represents a particular configuration. Thus, it is possible for the fielded system to be in any configuration and operating in any demand environment at any particular point in time. Since configuration is under the control of the decision-maker, he or she can choose to move to any vertical column of the spindle at any point in time (recognizing that it takes one time step to make this move). However, the demand environment is not under the control of the decision-maker. Illustrated in Fig. 14 is an instance where Config. 15 is operating in Demand Environment 1. If the decision-maker chooses to develop Config. 10 for the next time period, he or she is assured to move to the column corresponding to Config. 10[†††]; however, since the demand environment evolution is stochastic, the layer to which he or she moves is uncertain and depends on the evolution of the Markov chain specified by Step 2. Once the demand environment materializes, the decision-maker finds himself or herself at one particular total state and makes another decision about which of the 15 configurations to select for the following period.

Second, in order to apply the dynamic programming technique implied by Eq. (4), the multi-objective problem illustrated in Step 4 must be carefully converted to a single-objective problem. To do this, the present framework proposes to use the interpretation of the Pareto frontier as the set of optima for a weighted aggregate objective function over all possible weights. Thus, it is proposed that the Pareto frontier for the "closed-loop" case be found by forming an aggregate weighted objective function, solving the MDP problem as usual using this single objective, and repeating the process for a wide range of weights. While a simple additive weighting function is an appealing aggregate function, it suffers from an inability to detect concave segments of Pareto frontiers. To partially overcome this limitation, a heuristic technique using the variable-power per-period

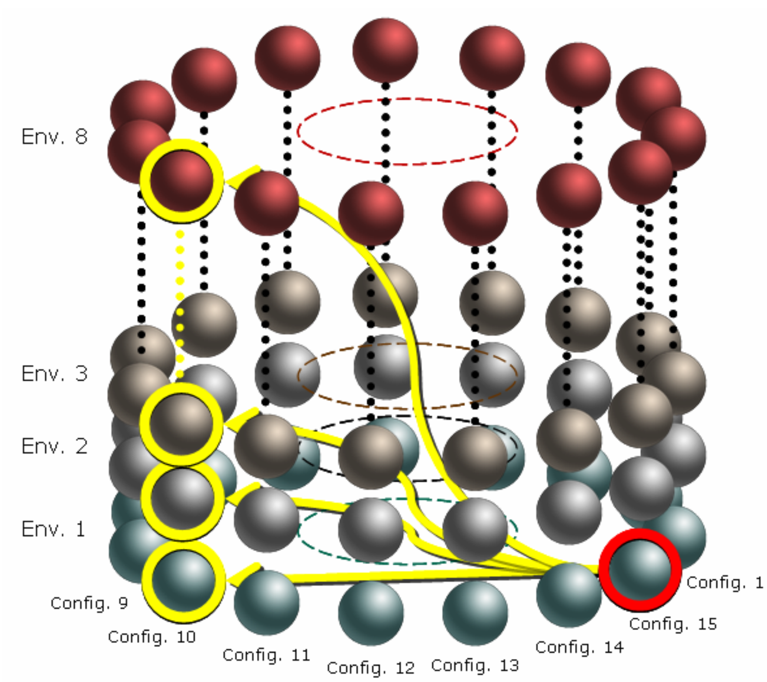

Figure 14. "Spindle" of Total States. *Each layer corresponds to one demand environment and each vertical column corresponds to one configuration. Environments 4-7 are not depicted. Arrows illustrate that, due to demand environment uncertainty, multiple possible total states are possible in the next period if a decision is made to transition from one configuration to another (e.g., Config. 15 to Config. 10).*

[†††] The assumption implicit in this assurance is that the decision-maker will not by accident develop a configuration other than Config. 10, which is likely to be reasonable in most cases.

aggregate objective function in Eq. (5) is used. In this equation, M is the number of per-period objectives, w_i is the weight on the i^{th} objective, T is the total number of time periods in the time horizon, y_i is per-period performance of the system in terms of the i^{th} objective (normalized such that the sum of y_i over all time periods cannot exceed unity or become negative, and such that higher values of y_i are preferred), and n is the objective function power.

$$h(s,a) = -\sum_{i=1}^{M} w_i\left(\tfrac{1}{T} - y_i(s,a)\right)^n \tag{5}$$

As a result of this formulation, which represents a natural adaptation of the problem framed in Steps 1-3, optimal solutions can be found efficiently for a range of decision-maker cost or performance preferences. These solutions take the form of a matrix with $|S|$ rows and T columns, where each element (s,t) indicates which of $|A|$ possible actions or decisions should be made given the system is in state s at time t. In other words, this matrix forms a *policy* by which the decision-maker should act to obtain optimal combinations of total cost and performance. In this example application, each policy matrix has dimensions 120 (states) × 5 (time periods), and 15 options exist for each element of the matrix. If a full-factorial analysis of all possible policies were to be conducted (as was done for the simple case of paths in Step 4A), $15^{600} = 10^{706}$ simulations would need to be executed! However, use of the structure of the problem as posed by Eq. (4) and scanning over weights and powers as suggested in Eq. (5) permits optimal policy solutions to be found within hours on a standard desktop computer.

Expected cost and performance results for policy solutions to the distributed-payload satellite system application are shown by each blue "x" in Fig. 15. Among these solutions, the nondominated (Pareto-optimal) solutions are highlighted and connected in red. Note that the minimum-cost and maximum-performance endpoints of the Pareto frontier are identical to those of the open-loop full factorial analysis of Fig. 12, and the shape of the frontier largely mirrors that of Fig. 12.[‡‡‡] However, an interesting solution with performance superior to any available from an open-loop path is visible at an expected total cost of $40 million. Depicted in Fig. 16 in the same format as the open-loop results earlier, it can be seen that this policy solution is nearly the same as the "do nothing" policy but with one exception: As the top left plot shows, at the t = 0 time period the policy occasionally (in 14.5% of cases) calls for a decision to develop and subsequently field the three-payload monolith. Whether decision is made is governed by the demand environment, as the policy indicates in Table 7. In this table, the policy solution itself is shown, and the action specified by the policy is provided for a system in any state s (the row) at any time t (the column). Looking only at the eight total states that are associated with Config. 1 (i.e., total states 1, 16, 31, 46, 61, 76, 91, and 106), it can be seen that the decision to develop Config. 11 rather than Config. 1 at t = 0 occurs only in total states 91 and 106, which correspond to a situation in which either the 2+3 or 1+2+3 demand environment

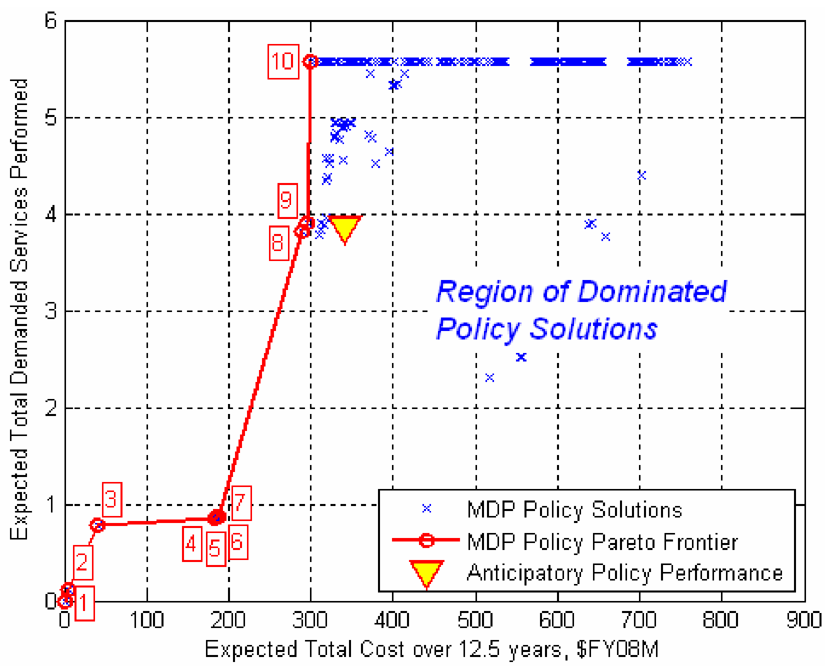

Figure 15. Trade between total demanded services performed and total cost for MDP policy solutions.

‡‡‡ The sparsity of points on this frontier is largely due its concavity: Only four of the frontier points could be found using n = 1 in Eq. (5). The heuristic method adopted for improving the frontier estimate by increasing n beyond unity was only partially successful in identifying the full frontier, and this is a clear area for future development.

exists. In other words, this policy achieves a low expected cost and an appreciable expected performance by only developing the three-payload monolith if a substantial demand for services materializes early during the program. Such a result is impossible to capture using the fixed configuration paths of Step 4A.

Figure 15 also permits comparisons to be made with policies that might be brainstormed or proposed outside of the MDP solution procedure. For example, one reasonable policy that might be proposed is to always develop and field the configuration that least expensively maximizes performance in the most likely next-period demand environment.[§§§] The policy implied by this statement is provided in Table 8; for instance, if Config. 2 (the PL1-only configuration) is currently operational in the "1+2" demand environment (i.e., if the system is in total state 62), the most likely next-period demand environment according to Table 5 is also the "1+2" demand. To least expensively fulfill both the PL1 and PL2 functions demanded in this environment, a single PL2-only module would be developed and launched, which places the system into Config. 8. Thus, as Table 8 shows, Config. 8 is the decision made from total state 62 at all except the final time period.[****]

The performance of this next-period anticipatory policy is summarized by the yellow triangle in Fig. 15 and detailed in Fig. 17. Figure 15 in particular illustrates two interesting and important points regarding this anticipatory policy: First, this policy is dominated by others discovered in the optimization process: Both policies 9 and 10 on the Pareto frontier perform, on average, more demanded services at a lower cost. Second, this anticipatory policy is just one of many options; even if it were nondominated, selection of this particular policy carries with it no options regarding cost and performance preferences. In contrast, a search throughout the policy design space (as was completed in order to produce Fig. 15) allows the decision-maker to understand the cost and performance trades available and select a policy according to his or her preferences.

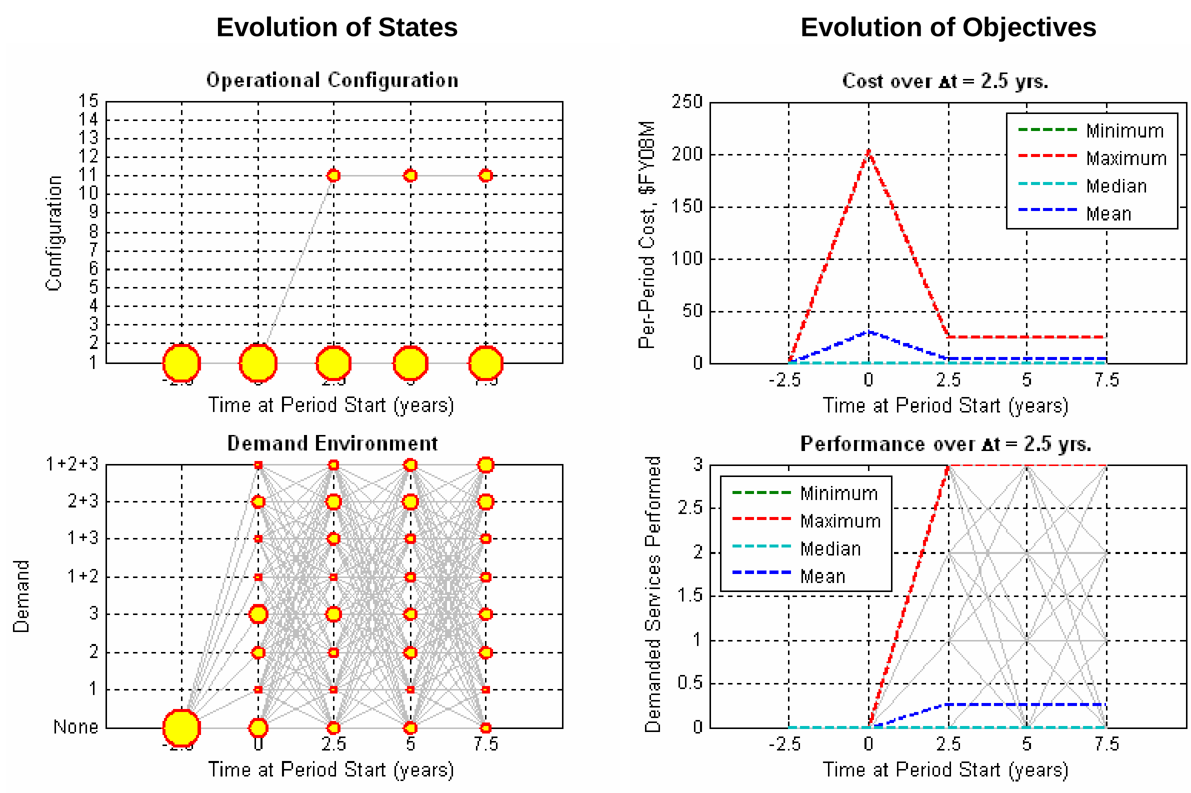

Figure 16. Evolution of states and performance for Pareto-optimal policy #3 (defined in Table 7). *In the plots on the left, the size of circles indicate the relative number of Monte Carlo simulation cases that exist in a given configuration or demand environment state (on the y-axes) at a given time (on the x-axes). The plots on the right indicate the associated evolution of per-period cost and performance. In all plots, gray lines indicate transitions made in at least one simulation.*

[§§§] In the event that multiple demand environments have the same probability of materializing next, the environment with the demand for more services is used.

[****] The reason for the difference in the final time period decision is the same as discussed earlier in Section IV.D.1.

Table 7. Pareto-optimal policy #3.

Configuration decisions for a system in state s *at time* t *are indicated by matrix elements shaded in gray.*

Current State, *s*			Time at Period Start (years), *t*				
Total State	Env.	Config.	-2.5	0	2.5	5	7.5
1	None	1	1	1	1	1	1
2	None	2	11	12	1	1	2
3	None	3	11	11	3	1	3
4	None	4	14	10	4	4	1
5	None	5	14	5	5	5	1
6	None	6	13	13	6	6	6
7	None	7	7	7	7	7	1
8	None	8	11	8	8	8	1
9	None	9	15	15	9	9	4
10	None	10	10	10	10	10	10
11	None	11	11	11	11	11	11
12	None	12	12	12	12	12	2
13	None	13	13	13	13	13	6
14	None	14	14	14	14	14	1
15	None	15	15	15	15	15	2
16	1	1	11	1	1	1	1
17	1	2	11	11	2	2	2
18	1	3	11	11	3	1	3
19	1	4	14	14	4	4	4
20	1	5	11	5	5	5	1
21	1	6	13	13	6	6	1
22	1	7	7	7	7	7	7
23	1	8	11	8	8	8	2
24	1	9	15	15	15	9	1
25	1	10	10	10	10	10	10
26	1	11	11	11	11	11	1
27	1	12	12	12	12	12	2
28	1	13	13	13	13	13	3
29	1	14	14	14	14	14	1
30	1	15	15	15	15	15	9
31	2	1	11	1	1	1	1
32	2	2	11	11	8	1	1
33	2	3	11	11	3	3	1
34	2	4	14	10	10	4	1
35	2	5	11	5	5	5	1
36	2	6	13	13	13	6	1
37	2	7	7	7	7	7	1
38	2	8	11	8	8	8	8
39	2	9	15	15	15	9	1
40	2	10	10	10	10	10	4
41	2	11	11	11	11	11	11
42	2	12	12	12	12	12	2
43	2	13	13	13	13	13	1
44	2	14	14	14	14	14	4
45	2	15	15	15	15	15	8
46	3	1	11	1	1	1	1
47	3	2	11	11	12	2	2
48	3	3	11	11	3	3	1
49	3	4	14	14	10	4	1
50	3	5	14	5	5	5	1
51	3	6	13	13	13	6	6
52	3	7	7	7	7	7	7
53	3	8	11	8	8	8	3
54	3	9	15	15	15	9	9
55	3	10	10	10	10	10	4
56	3	11	11	11	11	11	1
57	3	12	12	12	12	12	1
58	3	13	13	13	13	13	6
59	3	14	14	14	14	14	5
60	3	15	15	15	15	15	10
61	1+2	1	11	1	1	1	1
62	1+2	2	11	12	8	8	1
63	1+2	3	11	11	3	3	3
64	1+2	4	14	14	14	10	4
65	1+2	5	5	5	5	5	5
66	1+2	6	13	13	13	13	6
67	1+2	7	7	7	7	7	7
68	1+2	8	11	8	8	8	2
69	1+2	9	15	15	15	15	9
70	1+2	10	14	10	10	10	3
71	1+2	11	11	11	11	11	1
72	1+2	12	12	12	12	12	1
73	1+2	13	13	13	13	13	6
74	1+2	14	14	14	14	14	4
75	1+2	15	15	15	15	15	15
76	1+3	1	11	1	1	1	1
77	1+3	2	11	12	12	2	2
78	1+3	3	11	11	11	3	1
79	1+3	4	14	14	10	4	1
80	1+3	5	14	14	5	5	5
81	1+3	6	13	13	13	6	6
82	1+3	7	7	7	7	7	7
83	1+3	8	11	11	8	8	8
84	1+3	9	15	15	15	9	9
85	1+3	10	10	10	10	10	1
86	1+3	11	11	11	11	11	11
87	1+3	12	12	12	12	12	12
88	1+3	13	13	13	13	13	3
89	1+3	14	14	14	14	14	5
90	1+3	15	15	15	15	15	4
91	2+3	1	11	11	1	1	1
92	2+3	2	11	11	12	2	2
93	2+3	3	11	11	3	3	3
94	2+3	4	14	14	10	4	4
95	2+3	5	14	5	5	5	5
96	2+3	6	13	13	13	6	1
97	2+3	7	7	7	7	7	7
98	2+3	8	11	8	8	8	2
99	2+3	9	15	15	15	9	4
100	2+3	10	10	10	10	10	4
101	2+3	11	11	11	11	11	1
102	2+3	12	12	12	12	12	12
103	2+3	13	13	13	13	13	13
104	2+3	14	14	14	14	14	4
105	2+3	15	15	15	15	15	15
106	1+2+3	1	11	11	1	1	1
107	1+2+3	2	11	11	12	8	2
108	1+2+3	3	11	11	11	3	3
109	1+2+3	4	14	14	10	10	1
110	1+2+3	5	11	14	5	5	5
111	1+2+3	6	13	13	13	13	6
112	1+2+3	7	7	7	7	7	1
113	1+2+3	8	11	11	8	8	8
114	1+2+3	9	15	15	15	15	4
115	1+2+3	10	10	10	10	10	3
116	1+2+3	11	11	11	11	11	11
117	1+2+3	12	12	12	12	12	2
118	1+2+3	13	13	13	13	13	3
119	1+2+3	14	14	14	14	14	1
120	1+2+3	15	15	15	15	15	15

E. Step 5: Implications for Initial System Selection

Early in this paper it was emphasized that a major purpose of this framework is to inform initial system selection. The analysis of Step 4 has produced a large set of data on optimal paths and policies to follow for the *entire* system time horizon, and it is easy to lose track of the implications this has for the *initial* system decision. This final step of the framework builds upon the analysis results of Step 4 to provide implications for this decision.

1. Implications based on the Expected-Value Pareto Frontier

In the case of a path, the initial decision is simply the first configuration in its associated configuration sequence. In the case of a policy, the initial decision is found by locating the initial condition in the row of the policy matrix (in this distributed-payload satellite application, at total state 1, which corresponds to the "nothing" configuration fielded and no services demanded) and examining the element in the first column (in this case, the t = -2.5 year column). To facilitate this, the initial configurations specified by the Pareto-optimal paths and policies found in Figs. 12 and 15 are identified in Fig. 18. In this figure, the Pareto frontier solutions of Figs. 12 and 15 are identified by their expected total cost on the x-axis. On the y-axis are the initial configuration decisions called for by each Pareto-optimal path (yellow circles) or policy (blue squares). Two particular observations can be made: First, only three configurations (Configs. 1, 3, and 11) appear among the optimal initial decisions. All paths and policies with other initial decisions are dominated by paths and policies using these three configurations. Second, the size of the initial configuration tends to increase as the expected total cost of the system increases. For example, only the "Nothing" configuration (Config. 1) appears as an optimal initial decision for total expected budgets under \$195 million; these solutions tend to be either policies that wait until sufficient demand materializes to justify the expenditure of funds or paths that tend to delay initial operational capability until demand evolves substantially beyond the initial "None" environment. At the highest expected total cost is the decision to initially develop the three-payload monolith (Config. 11), which is the least expensive method to ensure complete capture of all possible future demand for services.

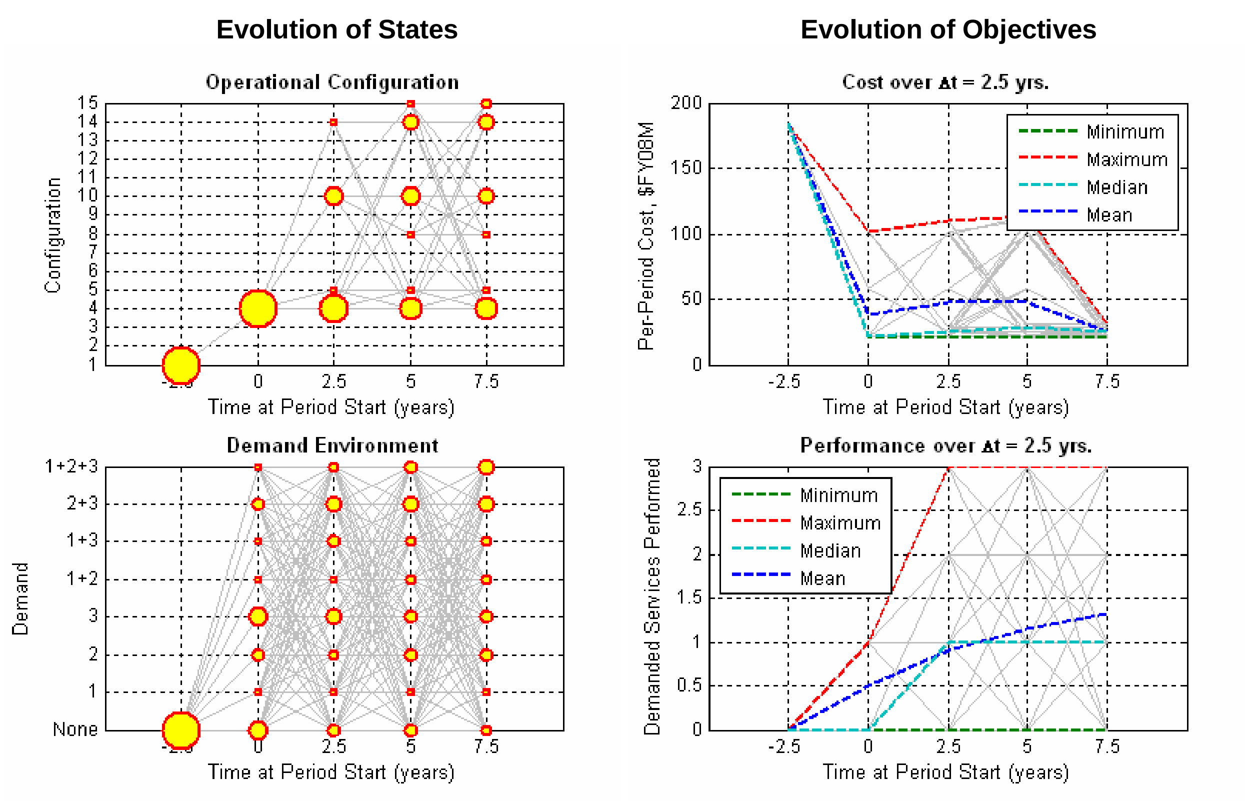

Figure 17. Evolution of states and performance for an anticipatory policy (defined in Table 8).
In the plots on the left, the size of circles indicate the relative number of Monte Carlo simulation cases that exist in a given configuration or demand environment state (on the y-axes) at a given time (on the x-axes). The plots on the right indicate the associated evolution of per-period cost and performance. In all plots, gray lines indicate transitions made in at least one simulation.

Table 8. Anticipatory Policy.

Configuration decisions for a system in state s *at time* t *are indicated by matrix elements shaded in gray.*

Current State, *s*			Time at Period Start (years), *t*				
Total State	**Env.**	**Config.**	**-2.5**	**0**	**2.5**	**5**	**7.5**
1	None	1	4	4	4	4	1
2	None	2	4	4	4	4	1
3	None	3	4	4	4	4	1
4	None	4	4	4	4	4	1
5	None	5	4	4	4	4	1
6	None	6	6	6	6	6	1
7	None	7	7	7	7	7	1
8	None	8	4	4	4	4	1
9	None	9	4	4	4	4	1
10	None	10	4	4	4	4	1
11	None	11	11	11	11	11	1
12	None	12	7	7	7	7	1
13	None	13	6	6	6	6	1
14	None	14	4	4	4	4	1
15	None	15	4	4	4	4	1
16	1	1	4	4	4	4	1
17	1	2	4	4	4	4	1
18	1	3	4	4	4	4	1
19	1	4	4	4	4	4	1
20	1	5	4	4	4	4	1
21	1	6	6	6	6	6	1
22	1	7	7	7	7	7	1
23	1	8	4	4	4	4	1
24	1	9	4	4	4	4	1
25	1	10	4	4	4	4	1
26	1	11	11	11	11	11	1
27	1	12	7	7	7	7	1
28	1	13	6	6	6	6	1
29	1	14	4	4	4	4	1
30	1	15	4	4	4	4	1
31	2	1	7	7	7	7	1
32	2	2	7	7	7	7	1
33	2	3	10	10	10	10	1
34	2	4	10	10	10	10	1
35	2	5	14	14	14	14	1
36	2	6	13	13	13	13	1
37	2	7	7	7	7	7	1
38	2	8	10	10	10	10	1
39	2	9	10	10	10	10	1
40	2	10	10	10	10	10	1
41	2	11	11	11	11	11	1
42	2	12	7	7	7	7	1
43	2	13	13	13	13	13	1
44	2	14	14	14	14	14	1
45	2	15	10	10	10	10	1
46	3	1	4	4	4	4	1
47	3	2	4	4	4	4	1
48	3	3	4	4	4	4	1
49	3	4	4	4	4	4	1
50	3	5	4	4	4	4	1
51	3	6	6	6	6	6	1
52	3	7	7	7	7	7	1
53	3	8	4	4	4	4	1
54	3	9	4	4	4	4	1
55	3	10	4	4	4	4	1
56	3	11	11	11	11	11	1
57	3	12	7	7	7	7	1
58	3	13	6	6	6	6	1
59	3	14	4	4	4	4	1
60	3	15	4	4	4	4	1
61	1+2	1	5	5	5	5	1
62	1+2	2	8	8	8	8	1
63	1+2	3	8	8	8	8	1
64	1+2	4	5	5	5	5	1
65	1+2	5	5	5	5	5	1
66	1+2	6	13	13	13	13	1
67	1+2	7	12	12	12	12	1
68	1+2	8	8	8	8	8	1
69	1+2	9	8	8	8	8	1
70	1+2	10	8	8	8	8	1
71	1+2	11	11	11	11	11	1
72	1+2	12	12	12	12	12	1
73	1+2	13	13	13	13	13	1
74	1+2	14	5	5	5	5	1
75	1+2	15	8	8	8	8	1
76	1+3	1	11	11	11	11	1
77	1+3	2	12	12	12	12	1
78	1+3	3	13	13	13	13	1
79	1+3	4	14	14	14	14	1
80	1+3	5	14	14	14	14	1
81	1+3	6	13	13	13	13	1
82	1+3	7	12	12	12	12	1
83	1+3	8	15	15	15	15	1
84	1+3	9	15	15	15	15	1
85	1+3	10	15	15	15	15	1
86	1+3	11	11	11	11	11	1
87	1+3	12	12	12	12	12	1
88	1+3	13	13	13	13	13	1
89	1+3	14	14	14	14	14	1
90	1+3	15	15	15	15	15	1
91	2+3	1	7	7	7	7	1
92	2+3	2	7	7	7	7	1
93	2+3	3	10	10	10	10	1
94	2+3	4	10	10	10	10	1
95	2+3	5	14	14	14	14	1
96	2+3	6	13	13	13	13	1
97	2+3	7	7	7	7	7	1
98	2+3	8	10	10	10	10	1
99	2+3	9	10	10	10	10	1
100	2+3	10	10	10	10	10	1
101	2+3	11	11	11	11	11	1
102	2+3	12	7	7	7	7	1
103	2+3	13	13	13	13	13	1
104	2+3	14	14	14	14	14	1
105	2+3	15	10	10	10	10	1
106	1+2+3	1	11	11	11	11	1
107	1+2+3	2	12	12	12	12	1
108	1+2+3	3	13	13	13	13	1
109	1+2+3	4	14	14	14	14	1
110	1+2+3	5	14	14	14	14	1
111	1+2+3	6	13	13	13	13	1
112	1+2+3	7	12	12	12	12	1
113	1+2+3	8	15	15	15	15	1
114	1+2+3	9	15	15	15	15	1
115	1+2+3	10	15	15	15	15	1
116	1+2+3	11	11	11	11	11	1
117	1+2+3	12	12	12	12	12	1
118	1+2+3	13	13	13	13	13	1
119	1+2+3	14	14	14	14	14	1
120	1+2+3	15	15	15	15	15	1

Also noted next to several paths and policies in Fig. 18 are the number of transitions Φ available from each initial configuration (1, 3, or 11) for the average per-period cost associated with each total cost. As discussed in Step 1, this number Φ is an indicator of flexibility, and it can be seen that more flexible initial configurations (Φ = 2 or Φ = 3) are selected at higher cost and performance preferences. Thus, there exists some correlation between flexibility and performance. However, the maximum-performance (and maximum-cost) Config. 11 initial decision is far from the most flexible for its average $60 million per-period budget; Fig. 9 illustrates that the fully-fractionated three-payload configuration (Config. 15) has Φ = 8 transitions available for the same budget. Thus, this example illustrates that maximization of performance does not necessarily translate into maximization of the flexibility of a system's configuration.

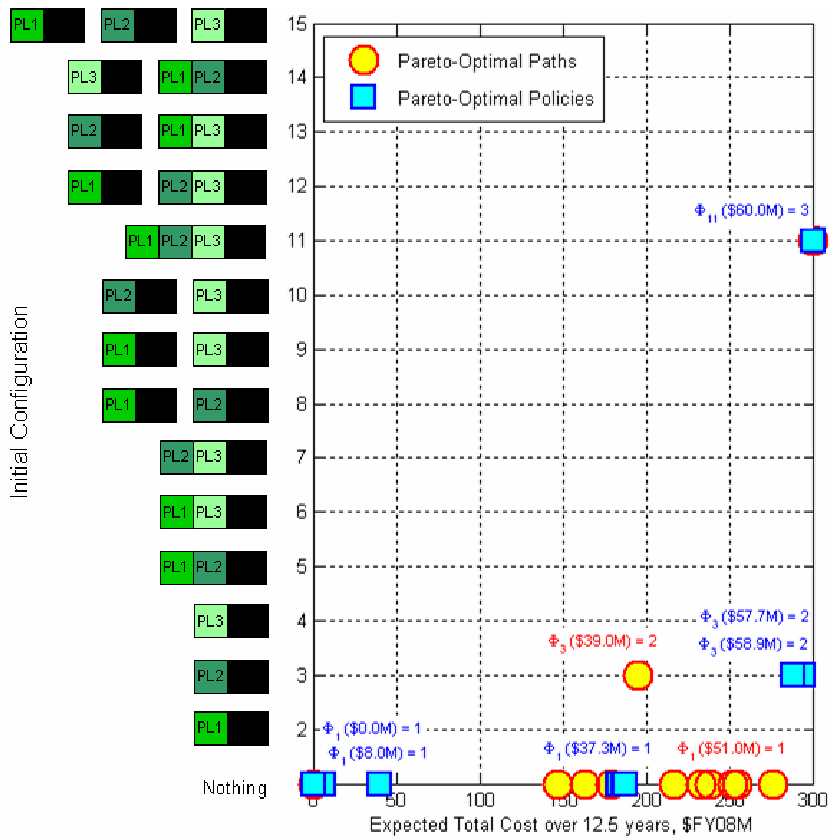

Figure 18. Initial configurations for Pareto-optimal paths and policies as a function of expected path or policy total cost. *Also noted are the number of transitions available for several initial configurations at their path or policy's average per-period budget requirements.*

2. Accounting for Non-Expected-Value Objectives

A final relevant consideration for initial system selection is the fact that expected-value objective functions for the cumulative cost and performance metrics may not fully capture a decision-maker's true objectives. Use of these expected-value objectives enables the use of MDP dynamic programming techniques to efficiently explore the astronomically large policy trade-space; however, in the case of one-of-a-kind satellite programs a decision-maker may also be interested in minimizing risks associated with a given expected level of cost or performance.

Operating under the assumption that the expected-value optima discovered in Step 4 are reasonable initial guesses for desirable policies, a specialized multi-objective genetic algorithm may be employed to perturb each of the policies identified in Fig. 15, simulate each new hybrid policy, and search for non-dominated solutions in terms of any combination of metrics that can be accounted for via simulation. The results of Fig. 19 are produced by applying this technique to the new metrics of 90th percentile (near-worst-case) total cost and 10th percentile (near-worst-case) total number of demanded services performed, in addition to the expected-value versions of these metrics. Of particular note in the Fig. 19 multivariate plot are four subplots: First, the data in the subplot of the second row and first column shows the familiar expected-value cost and performance trade, with slightly better Pareto frontier performance due to the genetic algorithm's search. Second, the data in the subplot of the last row and second column shows the 10th percentile performance vs. the 90th percentile cost; the performance data in this subplot is noticeably more discrete since fractional numbers of services performed are not possible in a simulation. Finally, the upper left and bottom right subplots show the correlations between the new percentile-based metrics and their expected-value counterparts. In the cases of both subplots, linear correlation is quite strong (R^2 = 0.85 and 0.88) and supports the use of expected value as a surrogate for optimizing the percentile-based metrics.

Also of note in Fig. 19 is that each data point, which represents a particular policy result, has a color that corresponds to the initial configuration decision implied by its associated policy. Of particular note is that these initial decisions differ little from those implied by the original path- and MDP-policy-based results in Fig. 18. Use of Config. 1 initially is still associated with low cost and performance; use of Config. 3 is associated with medium values for both objectives; and Config. 11 is associated with the highest levels of cost and performance. The primary difference is the introduction of Config. 13 as an initial decisions, which has performance and cost levels that are generally competitive with Config. 11.

The usefulness of the multivariate plot of Fig. 19 becomes more evident if cost or performance constraints are imposed by the decision-maker. For example, suppose that this decision-maker has a $500 million limit on the funds available for supporting this system over its time horizon. If the decision-maker wishes to be 90% sure that this budget will not be breached, a $500 million constraint may be imposed on the 90th percentile total cost metric. This constraint eliminates many high-cost (and also high-performance) options that formerly fell into the high 90th percentile cost regions of the multivariate plot that are now gray in Fig. 20. Similarly, the decision-maker may wish to have 90% confidence that more than one service will be performed over the system's lifetime. In this case an additional constraint may be imposed, represented by the horizontal gray stripe in the subplots of the last row in Fig. 20. Combined, these two constraints eliminate a large number of the policy options available. As Fig. 20, no policy options remain for which the "Nothing" configuration is acceptable. Furthermore, in both the expected-value-based and percentile-based performance vs. cost subplots, policies involving the three-payload monolith (Config. 11) as an initial configuration exhibit lower cost for the same (or better) performance as those that involve Config. 13. As a result, the decision is narrowed to one of whether to select a policy that suggests Config. 11 as an initial decision (at an expected and 90th percentile total cost of $300 million, with 5.6 expected services performed and 3 services performed in the 10th percentile) or, instead, Config. 3 (at an expected $285 million and 90th percentile $331 million total cost, with 3.8 expected services performed and 2 services performed in the 10th percentile). While no objectively correct decision exists, it is likely that the small ($15 million, or 5%) difference in expected cost and large (1.8 services, or 38%) difference in performance between the options would compel many decision-makers to accept the slightly higher budget for such a significant performance increase.

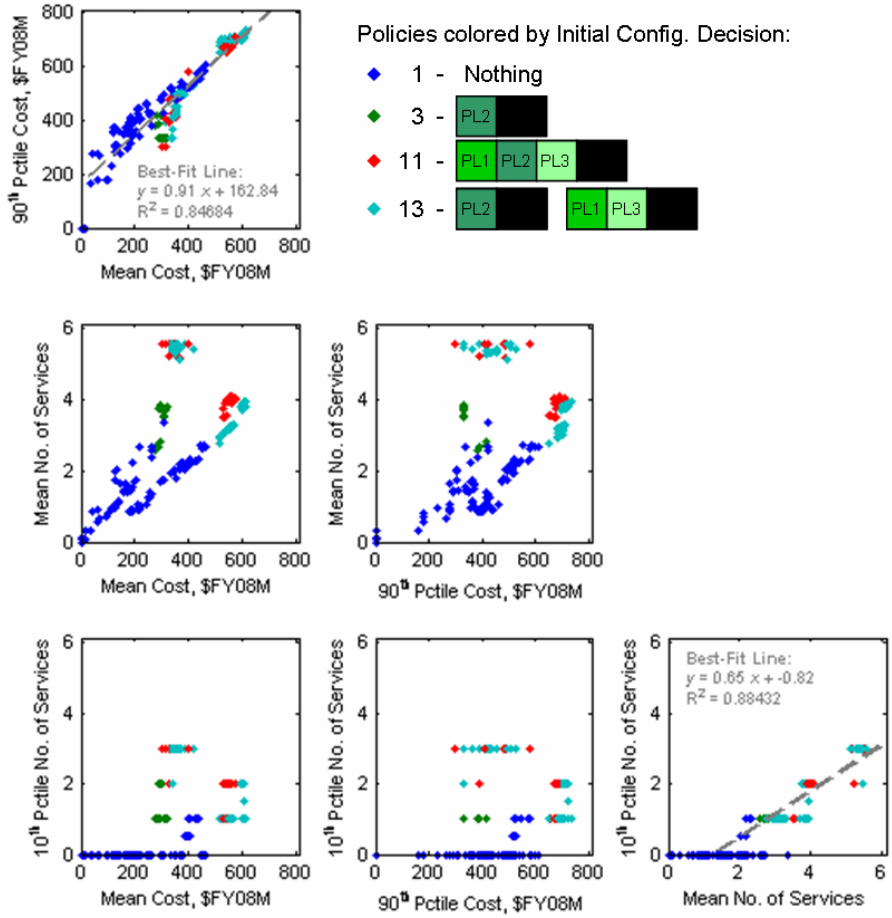

Figure 19. Multivariate plot of multi-objective genetic algorithm policy results. *Each data point indicates the performance of one policy result in terms of the four percentile-based and expected-value metrics of interest. Data points are colored by their corresponding policy's initial configuration decision.*

V. Summary and Implications

What conclusions can be drawn from the discussion over the last several pages? Ironically, the data and analysis presented for this distributed-payload satellite application largely identify paths and policies that involve an initial decision to develop a three-payload monolith – one of the *least* flexible options as identified in Step 1 – as the way to best respond to a changing demand environment at a low lifecycle cost. This is itself an interesting result, and it highlights the fact that integrating flexibility into space system design considerations is not synonymous with maximizing the flexibility of the system to be designed; rather, the benefits of flexibility must be traded against the costs. However, the contribution of this paper is intended to be much broader than this single case study result: As a consequence of this paper's Markovian state-space framework, flexibility can be quantitatively integrated into design decisions for a variety of space systems operating in a variety of potential environments. The present application serves as just one illustrative example. Future work will involve expansion of this framework to encompass a variety of additional effects and applications.

In short, the past decades have seen the state of the art in aerospace system design progress from a scope of simple optimization to one including robustness, with the objective of permitting a single system to perform well even in off-nominal future environments. Integrating flexibility – or the capability to easily modify a system after it has been fielded in response to changing environments – into system design represents a further step forward. One challenge in accomplishing this rests in that the decision-maker must consider not only the present system design decision, but also sequential future decisions. Despite widespread interest in the topic, the state of the art in designing flexibility into aerospace systems – and particularly space systems – tends to rely on analyses that are qualitative, deterministic, single-objective, and/or limited to consider only one future time period.

To address these gaps, the present work proposes a quantitative, stochastic, multi-objective, and multi-period framework for integrating flexibility into space system design decisions. Central to the framework are five steps. First, system configuration options are identified and costs of switching from one configuration to another are compiled into a cost transition matrix. Second, probabilities that demand on the system will transition from one mission to another are compiled into a mission demand Markov chain. Third, one performance matrix for each design objective is populated to describe how well the identified system configurations perform in each of the identified mission demand

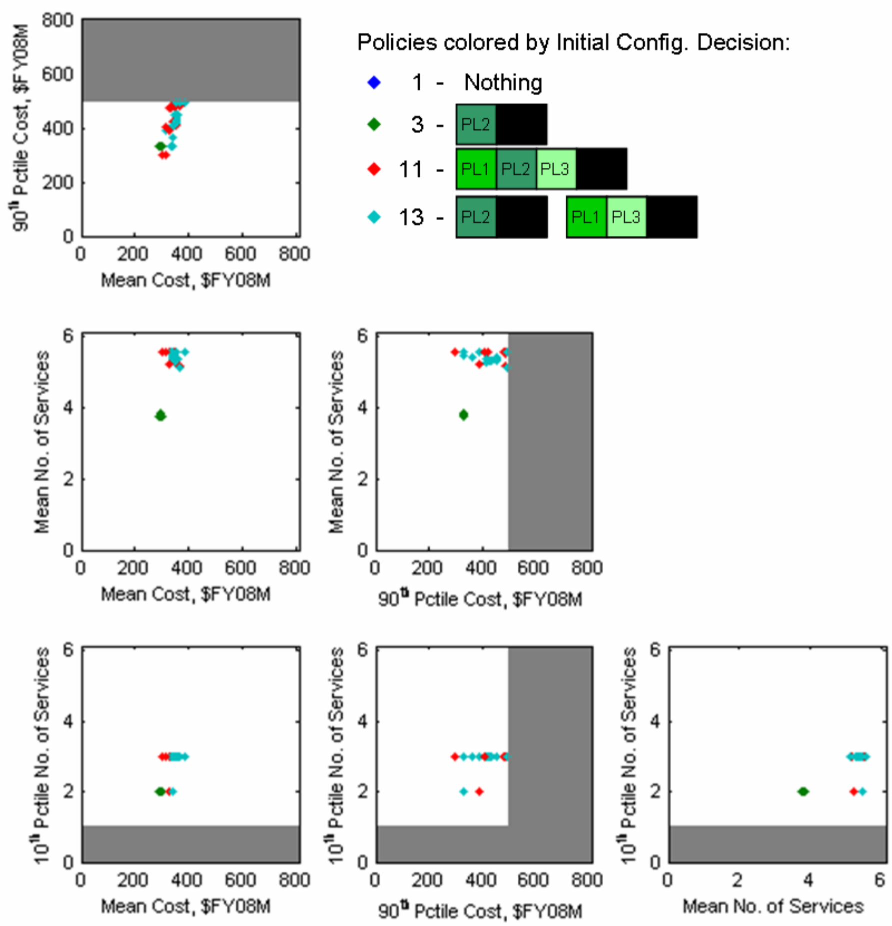

Figure 20. Multivariate plot of multi-objective genetic algorithm policy results with cost and performance constraints imposed. *Each data point indicates the performance of one policy result in terms of the four percentile-based and expected-value metrics of interest. Data points are colored by their corresponding policy's initial configuration decision. Gray areas indicate regions of the space eliminated due to cost and performance constraints.*

environments. Fourth, possible future sequences of system configurations are simulated and sequences that are Pareto-optimal in terms of the decision-maker's objectives are identified. In a complementary approach, the system decision problem is formulated as a multi-objective variant of a Markov decision process, and Pareto-optimal decision policies are identified. Finally, the paths and policies from the latter step are synthesized into a set of data to inform initial system selection.

The framework that this paper proposes builds on intuitive state-centric notions of flexibility from the previous economics and engineering literature and utilizes modeling and trade space exploration techniques from aerospace systems engineering and operations research to convert the flexibility problem into a comprehensive, tractable sequential decision-making problem. The result is a framework that is quantitative, stochastic, multi-objective, and multi-period in nature. In particular, the formulation of this problem posed here is amenable to solution through existing methods for Markov decision processes. Considering flexibility in this way enables the selection of systems *today*, tailored to the decision-maker's budget and preferences, that will be best able to perform when subject to a future of changing environments and requirements. It is hoped that the theoretical and practical contributions made through the work in this paper not only advance current thought on flexibility in the aerospace engineering literature, but also provide new and advanced tools to allow the space systems engineer to better design the vehicles and architectures that allow for the most effective exploration, utilization, and protection of the final frontier.

Acknowledgments

The author owes tremendous thanks to Dr. Joseph Saleh in the Space Systems Design Laboratory at the Georgia Institute of Technology for his invaluable advice and guidance throughout the development of this work. In addition, thanks are due to Joy Brathwaite, Jean-François Castet, Bradley Steinfeldt, Michael Grant, Gregory Lantoine, and George Burdell for the time that they willingly (or at times, perhaps reluctantly) sacrificed as sounding boards for ideas during the development of this framework and example. Finally, the author would like to acknowledge the support of the National Science Foundation through its Graduate Research Fellowship Program.

References

[1]"Broad Agency Announcement: System F6," DARPA Tactical Technology Office, BAA07-31, 16 July 2007, URL: http://www.darpa.mil/TTO/solicit/BAA07-31/F6_BAA_Final_07-16-07.doc [21 Nov. 2008].

[2]Wilson, J.R., "Satellite System F6: Divide and Conquer," *Aerospace America*, Vol. 47, No. 2, 2009, pp. 30-37.

[3]Brown, O.C., Eremenko, P., and Collopy, P.D., "Value-Centric Design Methodologies for Fractionated Spacecraft: Progress Summary from Phase 1 of the DARPA System F6 Program," AIAA 2009-6540, AIAA Space 2009 Conference and Exposition, Pasadena, 14-17 Sept. 2009.

[4]Augustine, N.R., Austin, W.M., Chyba, C.C., et. al., *Seeking a Human Spaceflight Program Worthy of a Great Nation*, Review of U.S. Human Spaceflight Plans Committee, Washington, Oct. 2009.

[5]Mish, F.C., "Flexible," *Merriam-Webster Collegiate Dictionary*, 11th ed., Merriam-Webster, Springfield, 2003, pp. 478-479.

[6]Saleh, J.H., Hastings, D.E., and Newman, D.J., "Flexibility in system design and implications for aerospace systems," *Acta Astronautica*, Vol. 53, No. 12, Dec. 2003, pp. 927-944.

[7]Knight, F.H., "Cost of Production and Price over Long and Short Periods," *The Journal of Political Economy*, Vol. 29, No. 4, April 1921, pp. 304-335.

[8]Hart, A.G., "Anticipations, Business Planning, and the Cycle," *The Quarterly Journal of Economics*, Vol. 51, No. 2, Feb. 1937, pp. 237-297.

[9]Stigler, G., "Production and Distribution in the Short Run," *The Journal of Political Economy*, Vol. 47, No. 3, June 1939, pp. 305-327.

[10]Koopmans, T.C., "On flexibility of future preference," *Human Judgments and Optimality*, Wiley, New York, 1964, pp. 243-254.

[11]Raiffa, H., *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, Addison-Wesley, Reading, 1968.

[12]Howard, R.A., "Decision Analysis: Applied Decision Theory," *Proceedings of the Fourth International Conference on Operational Research*, Wiley-Interscience, New York, 1966, pp. 55-71.

[13]Edwards, W., Miles, R.F., Jr., and Von Winterfeldt, D., *Advances in Decision Analysis: From Foundations to Applications*, Cambridge University Press, Cambridge, 2007.

[14]Schlaifer, R., *Probability and Statistics for Business Decisions*, McGraw-Hill, New York, 1959.

[15]Raiffa, H. and Schlaifer, R., *Applied Statistical Decision Theory*, Colonial Press, Clinton, 1961.

[16]Pratt, J.W., Raiffa, H., and Schlaifer, R., *Introduction to Statistical Decision Theory*, McGraw-Hill, New York, 1965.

[17]Daniel, W.W., *Decision Trees for Management Decision Making: An Annotated Bibliography*, Public Administration Series Bibliography P-254, Vance Bibliographies, Monticello, 1979.

[18]Oh, C.R., *Evaluating Flexibility in Railroad Construction Projects*, Master's Thesis, Massachusetts Institute of Technology, June 2005.

[19]Kirkwood, C.W., *Decision Tree Primer*, Arizona State University, Tempe, 2002, URL: http://www.public.asu.edu/~kirkwood/DAStuff/decisiontrees/index.html [2 Oct. 2010].
[20]Saleh, J.H., Lamassoure, E.S., Hastings, D.E, and Newman, D.J., "Flexibility and the Value of On-Orbit Servicing: New Customer-Centric Perspective," *Journal of Spacecraft and Rockets*, Vol. 40, No. 2, March 2003, pp. 279-291.
[21]Saleh, J.H., *Weaving Time into System Architecture: New Perspectives on Flexibility, Spacecraft Design Lifetime, and On-Orbit Servicing*, Ph.D. Thesis, Massachusetts Institute of Technology, Feb. 2002.
[22]Saleh, J.H., Mark, G., and Jordan, N.C., "Flexibility: A Multi-Disciplinary Literature Review and a Research Agenda for Designing Flexible Engineering Systems," *Journal of Engineering Design*, Vol. 20, No. 3, June 2009, pp. 307-323.
[23]Jones, R.A. and Ostroy, J.M., "Flexibility and Uncertainty," *Rev. of Economic Studies*, Vol. 51, No. 1, Jan. 1984, pp. 13-32.
[24]Christian, J.A. III, "A Quantitative Approach to Assessing System Evolvability," NASA Johnson Space Center, 2004.
[25]Christian, J.A. III and Olds, J.R., "A Quantitative Methodology for Identifying Evolvable Space Systems," AIAA 2005-2543, 1st Space Exploration Conference: Continuing the Voyage of Discovery, Orlando, 30 Jan. – 1 Feb. 2005.
[26]Gupta, S.K. and Rosenhead, J., "Robustness in Sequential Investment Decisions," *Management Science*, Vol. 15, No. 2, Oct. 1968, pp. B18-B29.
[27]Baykasoğlu, A., "Quantifying machine flexibility," *International Journal of Production Research*, Vol. 47, No. 15, Aug. 2009, pp. 4109-4123.
[28]Silver, M.R. and de Weck, O.L., "Time-Expanded Decision Network: A New Framework for Designing Evolvable Complex Systems," AIAA 2006-6964, 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Portsmouth, 6-8 Sept. 2006.
[29]Silver, M.R. and de Weck, O.L., "Time-Expanded Decision Networks: A Framework for Designing Evolvable Complex Systems," *Systems Engineering,* Vol. 10, No. 2, April 2007, pp. 167-186.
[30]M. Mandelbaum and J. Buzacott, "Flexibility and decision making," European Journal of Operational Research, Vol. 44, No. 1, Jan. 1990, pp. 17-27.
[31]Viscito, L. and Ross, A.M., "Quantifying Flexibility in Tradespace Exploration: Value Weighted Filtered Outdegree," AIAA 2009-6561, Space 2009 Conference and Exposition, Pasadena, 14-17 Sept. 2009.
[32]Ross, A.M. and Rhodes, D.H., "Using Natural Value-Centric Time Scales for Conceptualizing System Timelines through Epoch-Era Analysis," INCOSE International Symposium 2008, Utrecht, June 2008.
[33]Mark, G.T., *Incorporating Flexibility into System Design: A Novel Framework and Illustrated Developments*, M.S. Thesis, Massachusetts Institute of Technology, June 2005.
[34]Nilchiani, R., *Measuring Space Systems Flexibility: A Comprehensive Six-Element Framework*, Ph.D. Thesis, Massachusetts Institute of Technology, Sept. 2005.
[35]Lim, D., *A Systematic Approach to Design for Lifelong Aircraft Evolution*, Ph.D. Thesis, Georgia Institute of Technology, May 2009.
[36]Lim, D.W. and Mavris, D.N., "A Design Methodology for Lifelong Aircraft Evolution," AIAA 2010-0283, 48th Aerospace Sciences Meeting, Orlando, 4-7 Jan. 2010.
[37]Stanley, D., Cook, S., Connolly, J., et. al., *NASA's Exploration Systems Architecture Study: Final Report*, NASA-TM-2005-214062, Nov. 2005, Preface, Chap. 6.
[38]Bernanke, B.S., "Irreversibility, Uncertainty, and Cyclical Investment," *The Quarterly Journal of Economics*, Vol. 98, No. 1, Feb. 1983, pp. 85-106.
[39]Lafleur, J.M. and Saleh, J.H., "Exploring the F6 Fractionated Spacecraft Trade Space with GT-FAST," AIAA 2009-6802, AIAA Space 2009 Conference and Exposition, Pasadena, 14-17 Sept. 2009.
[40]Eremenko, P., "System F6," DARPA Tactical Technology Office Presentation, Feb. 2010, URL: http://www.darpa.mil/WorkArea/DownloadAsset.aspx?id=2648 [28 June 2011].
[41]Brown, O., "Program of Record Analysis." DARPA/TTO Memorandum, 6 May 2008.
[42]Larson, W.J., and Wertz, J.R. (ed.), *Space Mission Analysis and Design*, 3rd ed., Microcosm Press and Kluwer Academic Publishers, El Segundo, 1999.
[43]NASA Goddard Space Flight Center, *National Space Science Data Center Spacecraft Query* [online database]. URL: http://nssdc.gsfc.nasa.gov/nmc/SpacecraftQuery.jsp [13 Feb. 2009].
[44]Martin, D.H., Anderson, P.R., and Bartamian, L., *Communication Satellites*, 5th ed., The Aerospace Press, El Segundo, 2007.
[45]"SSTL VHRI 250," Surrey Satellite Technology, URL: http://www.sst-us.com/Downloads/Datasheets/Datasheet_VHRI250_OL [6 Aug. 2011].
[46]Carrel, A.R., Cawthorne, A.D., Richardson, G., and Gomes, L.M., "NigeriaSat-2: Advances in GN&C that Enable a High Performance Small Satellite Earth Observation Mission," AAS 10-033, 33rd AAS Rocket Mountain Guidance and Control Conference, Breckenridge, 5-10 Feb. 2010.
[47]Lafleur, J.M. and Saleh, J.H., "GT-FAST: A Point Design Tool for Rapid Fractionated Spacecraft Sizing and Synthesis," AIAA 2009-6563, AIAA Space 2009 Conference and Exposition, Pasadena, 14-17 Sept. 2009.
[48]NASA Johnson Space Center, Mission Operations Cost Model, Online Software, 25 May 2007, URL: http://cost.jsc.nasa.gov/MOCM.html [25 July 2011].
[49]Winston, W.L., *Operations Research: Applications and Algorithms*, 4th ed., Duxbury Press, Pacific Grove, 2003, chap. 11.
[50]Sen, P. and Yang, J., *Multiple Criteria Decision Support in Engineering Design*, Springer, London, 1998, chap. 4.