The Pennsylvania State University

The Graduate School

Department of Aerospace Engineering

GENERAL FRAMEWORK FOR THE RECONFIGURATION OF SATELLITE CONSTELLATIONS

A Dissertation in

Aerospace Engineering

by

Matthew Phillip Ferringer

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2009

The dissertation of Matthew Phillip Ferringer was reviewed and approved* by the following:

David Spencer
Associate Professor of Aerospace Engineering
Dissertation Advisor
Chair of Committee

Robert Melton
Professor of Aerospace Engineering

Sven Bilén
Associate Professor of Engineering Design, Electrical Engineering, and Aerospace Engineering

Patrick Reed
Associate Professor of Civil and Environmental Engineering

Thomas Starchville Jr.
Director of Mission Analysis and Operations Department, The Aerospace Corporation
Special Member

George Lesieutre
Professor of Aerospace Engineering
Head of the Department of Aerospace Engineering

*Signatures are on file in the Graduate School

ABSTRACT

From remote sensing to navigation and communication, satellite constellations have become an indispensible component of our society's infrastructure. Recent events, including China's intercept of their Feng Yun-1C weather satellite and the United States' intercept of a non-functioning satellite, have dramatically increased the amount of space debris, which poses an increased risk for on-orbit collisions. When the loss or degradation of a satellite in a constellation is experienced – be it from a collision with space debris, on-orbit malfunctions, or natural causes – the constellation may no longer be capable of fulfilling its mission requirements. Instead of simply accepting the degraded performance, stakeholders may consider reconfiguration of the remaining spacecraft. In this research, a general framework for the reconfiguration of satellite constellations is developed.

The key characteristic that separates this research from others that have come before it is that the future state of the reconfigured constellation is not assumed *a priori*; rather, it is this state that is found. In other words, the reconfigured geometry can occupy any feasible region (this is mission dependent) of the design space in an attempt to provide optimal performance with respect to multiple-criterion. These include the propellant expended, time of transfer, reduction in mission life, coverage performance, and risk due to maneuvering. When modeled as mathematical functions, some of these concerns exhibit continuous behavior; however, most have nonlinear, discrete, discontinuous, and/or multimodal characteristics.

The framework adapts a best-in-class parallel Multi-Objective Evolutionary Algorithm to approximate the optimal hypervolumes for this complex tradeoff-space. Several loss scenarios for the Global Positioning System constellation are presented to demonstrate the framework. An *a posteriori* procedure for decision support is introduced that enables down-selection to a final design from thousands of non-dominated reconfiguration alternatives.

Among the significant results to emerge from this research are the lessons learned as a result of the application of stochastic optimization to the constellation design problem domain. One such lesson indicates that objective functions related to coverage have the greatest influence on the multi-modality of the design space. Other results demonstrate that increasing the number of design variables and/or the application of operational constraints (such as fuel budgets) do not necessarily make the reconfiguration problem more difficult to solve; in some cases it becomes easier. The method itself is successful in providing a global context to decision makers that allows for defendable design selection in what was previously a computationally intractable optimization problem.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

No work such as this would have been possible without support of many. I would like to express my appreciation to members of The Aerospace Corporation's technical staff. Conversations with Ken Chan, Marc DiPrinzio, Rob Markin, Eric Campbell, Roger Thompson, Tom Gurlitz, Matthew Jones, and Scott Zimmer were invaluable in moving the research forward. I would especially like to express my gratitude to Tim Thompson and Ron Clifton. There are few better mentors than these gentlemen and I was privileged to have them has teammates and colleagues for this research.

To my advisor, David Spencer, who through some very challenging times, stuck by my side to develop a potential that I couldn't see. To my other committee members, Bob Melton, Pat Reed, Tom Starchville, and Sven Bilén for their knowledge and guidance over the years.

To Joshua Kollat for sharing his innovative multi-dimensional visualization software. To Penn State's Applied Research Labs (ARL) for the use of their tool, ARL Trade Space Visualizer.

To my family and especially my wife, Renee; they have all sacrificed, loved, encouraged, and inspired during this journey. This is for you.

Chapter 1

Introduction

In 1945, science fiction author and futurist, Arthur C. Clark wrote an article [1] for Wireless World magazine where he notes:

> "It will be observed that one orbit, with a radius of 42,000 km, has a period of exactly 24 hours. A body in such an orbit, if its plane coincided with that of the earth's equator, would revolve with the earth and would thus be stationary above the same spot on the planet. It would remain fixed in the sky of a whole hemisphere and unlike all other heavenly bodies would neither rise nor set… A single station could only provide coverage to half the globe, and for a world service three would be required, though more could be readily utilized."

Twelve years before the Soviet Union's launch of *Sputnik* 1, the world's first artificial satellite, Clark predicted the future. While some of the technical details in his article are not precise, Clark basically got it right. The geostationary orbit he described and subsequent suggestion of three "stations" distributed equally around the globe would indeed "be invaluable, not to say indispensable, in a world society."

The first satellite constellations, also referred to in literature as satellite arrays [2], consisted of single spacecraft typically with unique and independent missions. As performance requirements grew and technology evolved, multiple-spacecraft constellations (hereafter referred to as constellations) emerged. Though the applications vary widely from remote sensing, to navigation and communications, both entrepreneurs and governments alike spend billions of dollars annually to develop, build, test, and launch these complex systems. The considerable initial investment does not generate

returns for the stakeholders until the spacecraft are inserted into their operational orbits and begin performing their mission(s).

The operational lifetime of a constellation is threatened by a variety of phenomena, both man-made and natural, that will reduce the capacity and jeopardize its ability to perform its mission. The space debris environment is one such phenomenon and has produced several recorded collisions since the dawn of the space age [3]. In 1991 a defunct Russian navigation satellite (*Cosmos* 1934) collided with a fragment from another *Cosmos*. The next collision occurred in 1996 when an *Ariane* rocket stage hit one of their military satellites (*Cerise*). The final documented event, occurring in 2005 between a fragment from a Chinese rocket and United States rocket stage, produced three pieces of debris. Of these collisions, the only one affecting an operational satellite was the 1996 *Ariane/Cerise* and in every case, only several pieces of new debris were created. While these events are rare, the debris environment continues to grow. On January 11, 2007, China launched a ground-based medium-range ballistic missile and destroyed their *Feng Yun*-1C weather satellite orbiting 535 miles above the Earth. Just one year later, the United Sates intercepted a non-functioning satellite at an altitude of approximately 150 miles [4] adding to the over 12,000 pieces of currently tracked debris. The Chinese intercept resulted in the worst debris-producing collision in history [5] because the resulting pieces stay aloft much longer than those created by the low-altitude United States' mission. In fact, the U.S. mission was conducted to minimize the impact on the debris environment with most of the tracked fragments reentering within forty days.[1]

---

[1] http://www.space.com/news/080221-sat-shoot-spot.html, cited on October 7, 2008

Other phenomena that result in the loss of spacecraft include on-orbit malfunctions [6] and natural causes [7]. When a loss in capacity has occurred, for whatever reason, and mission performance must be restored quickly, it is typically not practical to build and/or launch spares. Rather, Decision Makers (DMs) can choose to reconfigure the remaining satellites such that the new constellation restores up to full satisfaction of mission requirements.

Previous work [8] [9] dealing with the reconfiguration problem has principally focused on minimizing the cost to transfer an initial configuration, constellation A, to a final configuration constellation B, both known *a priori*. Other work [10] presented a methodology for determining the optimal initial constellation that would allow for the lowest reconfiguration cost to a discrete set of future higher-capacity constellations. Both Scialom [8] and Siddiqi [10] dealt with global-communications constellations, so their use of well established Walker [11] or Streets-Of-Coverage (SOC) [12] design analysis techniques, to constrain the geometries for constellation B, was reasonable. However, when a constellation suffers a loss, the global coverage requirement may no longer be achievable with *any* reconfiguration of the remaining satellites. In this case, or when regional coverage is desired, these traditional constellation design methods are no longer appropriate. Another potential limitation observed in the literature [9] is the tendency to combine multiple objectives of the reconfiguration problem into a single cost function. Reducing a multiple-objective optimization problem to a single function requires the introduction of preference information *prior* to performing the optimization. The preference information has the potential to bias or constrain search such that an implicit

understanding of the design space tradeoffs are not easily obtained. These observed limitations with previous work motivate the goals of this thesis.

The primary goal of this research is to provide a *general* framework for the reconfiguration of satellite constellations. It will expand on the previous efforts by addressing the operational scenario where an existing constellation A has lost capacity for whatever reason, resulting in the need to optimally reconfigure to an undetermined constellation B. This fundamental departure from previous work, where constellation B is not known *a priori* or constrained to the discrete options offered by Walker or SOC constellations, requires the reconfiguration problem to be cast in a more general manner. In this work, the problem is posed using a multi-objective formulation wherein constellation B is designed such that performance is maximized at minimal reconfiguration cost and risk. Recent work [13] [14] [15] has demonstrated that parallel Multi-Objective Evolutionary Algorithms (pMOEAs) provide an efficient and effective means for solving constellation design optimization problems where traditional analysis methods fail. In this work, a pMOEA is used to solve the reconfiguration problem in a holistic manner wherein performance, cost, and risk objectives are optimized simultaneously during a single simulation. No individual optimal solution exists. Rather, a set that approximates the Pareto-optimal designs are found that represent the tradeoffs implicit to the reconfiguration problem.

Two secondary goals of this work are to demonstrate a procedure for decision support when choices must be made among high-dimensional tradeoffs; and to develop the lessons learned as a result of the application of stochastic optimization to the constellation design problem domain. The mathematical definition of optimization, taken

quite literally, means finding the best solution. While this conventional view of optimization has its place, the one-dimensional all-or-nothing approach is not adopted in this work. Rather a philosophy is embraced where the emphasis is to find new and innovative design principles by means of optimization techniques [16]. The resulting deeper understanding of the problem that emerges promotes innovation and discovery through visualization-based decision support.

The remainder of this thesis is organized as follows. Chapter 2 provides the background necessary to understand the satellite reconfiguration problem. Topics in this chapter include astrodynamic fundamentals, satellite constellation design, and satellite maneuvering. In Chapter 3, the confluence of multi-objective optimization, evolutionary algorithms, and parallel computing provide the underlying tools exploited by the general framework used in this thesis. Several reconfiguration case studies for the Global Positioning System (GPS) are detailed in Chapter 4. There, the mission-specific objective function formulations, constraints, evolutionary algorithm parameters, and parallel computing cluster configuration are described. In Chapter 5, the resulting tradeoff hypervolumes are presented and used to demonstrate the decision support procedure. For each case-study, the process of explicitly understanding the design space tradeoffs is shown and down-selection to the final configuration for constellation B, among the non-dominated hypervolume alternatives, is presented. The conclusions in the final chapter provide philosophical reflections on innovation and suggestions for future contributions made possible by adapting the general reconfiguration framework.

Chapter 2

Satellite Reconfiguration Background

In this chapter the fundamentals of Astrodynamics are provided with the key assumptions made to derive the Equations Of Motion (EOM) that give a satellite's position in its orbit at any time. Multiple satellites working together form a constellation. This chapter also provides a discussion of how a constellation's coverage is evaluated followed by other principal concerns of satellite reconfiguration including maneuvering, transfer time, life expectancy, and risk.

2.1 Astrodynamics Fundamentals

Astrodynamics refers to the study of the motion of objects in space subject to both natural and artificially induced forces. Vallado [17] notes that "true" astrodynamics has only existed since the 1950s and is the result of centuries of previous work in astronomy, ancient astrology, and celestial and orbital mechanics. Several of the most notable of early contributors are the duo of Johannes Kepler and Tycho Brahe.

2.1.1 On the Shoulders of Giants

Tycho Brahe (1546–1601), a Danish aristocrat, was fascinated by astronomy but frustrated with the accuracy of the tables of planetary motion. He dedicated his life and significant resources to building large astronomical instruments that enabled him to take

the most accurate observations of heavenly bodies of his time. Tycho believed the model of orbital motion of the planets was a hybrid between the Ptolemaic and Copernican systems [27]. In his view, the Earth was at rest while the remainder of the planets orbited the sun that was orbiting the Earth. Kepler, on the other hand, believed in the Copernican system wherein all planets orbited the sun while the sun remained "fixed". Kepler, aspiring to be a theologian, strongly believed that God must have had a geometric reason for placing the six planets (referring to the Moon, Mercury, Venus, Mars, Jupiter, and Saturn) at particular distances from the sun. Kepler became fixated on the idea that the planets were carried around by being embedded in rotating spheres. Given the uncertainties in the observations of the time, Kepler reasoned that his view might be the right one, but it wasn't until he used Brahe's observations, after Brahe's death in 1601, that Kepler was forced to abandon his geometric scheme to develop his three laws of planetary motion.

While Kepler's second law is referenced in subsequent discussions of orbital motion, all three [18] are noted here because of their importance.

1. The orbit of each planet is an ellipse with the Sun at one focus.

2. The line joining the planet to the Sun sweeps out equal areas in equal times.

3. The square of the period of a planet is proportional to the cube of its mean distance to the Sun.

These laws captured the kinematics of planetary motion, but the keys to modeling the dynamics were not uncovered until the late seventeenth century when Isaac Newton introduced his laws of motion and gravitation in Book I of *Principia* [19]. His second

law and his universal law of gravitation form the starting points for the derivation of the two-body EOM.

> Newton's second law of motion: The change of motion is proportional to the motive force impressed and is made in the direction of the right line in which that force is impressed.
>
> Newton's universal law of gravitation: Any two point masses attract each other with a force proportional to the product of their masses and inversely proportional to the square of the distance between them.

2.1.2 The Two-Body Problem

A satellite orbiting the Earth represents a two-body system with EOM that are found starting with Newton's second law. The proceeding derivation follows Vallado [17]. Newton's original language for his second law can be restated as,

$$\sum \vec{F} = \frac{d}{dt}\left(m\dot{\vec{r}}\right), \tag{2.1}$$

where to the force acting on a body, $\vec{F}$, is equal to the mass, *m*, times the inertial acceleration, $\ddot{\vec{r}}$, of that body. The geometry for the Earth–satellite system is shown in Fig. 2-1 to aid with the remainder of the derivation. The position vector, $\vec{r}$, locates the satellite relative to the $\hat{I}\hat{J}\hat{K}$ coordinate system, which neither rotates nor accelerates relative to the inertial $\hat{X}\hat{Y}\hat{Z}$ system. The $\hat{I}\hat{J}\hat{K}$ system is a geocentric inertial reference frame hereafter referred to as Earth Centered Inertial (ECI) where $\hat{I}$ points towards the

vernal equinox; the $\hat{J}$ axis is 90º to the East in the equatorial plane; and $\hat{K}$ extends through the North pole. Both the mass of the Earth, m_e, and satellite, m_{sat} are represented by points located $\vec{r}_e$ and $\vec{r}_{\text{sat}}$ relative to the $\hat{X}\hat{Y}\hat{Z}$ frame.

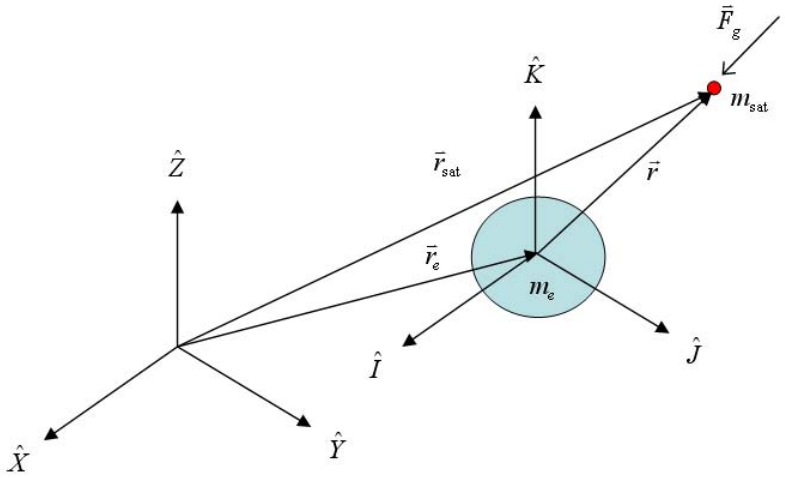


Figure 2-1: Two-body, Earth–satellite system in an inertial reference frame, adapted from Vallado [17]

Newton's law of gravitation is defined as,

$$\vec{F}_g = -\frac{Gm_e m_{\text{sat}}}{r^2}\left(\frac{\vec{r}}{r}\right), \tag{2.2}$$

where the force of gravity of the Earth, $\vec{F}_g$, acting on the satellite is equal to the product of their masses, the universal gravitational constant, G, and the inverse of the square of

the distance between them. The force of gravity acting on the Earth due to the satellite is identical but in the opposite direction. Substituting Eq. 2.1 into Eq. 2.2,

$$m_{\text{sat}}\ddot{\vec{r}}_{\text{sat}} = -\frac{Gm_e m_{\text{sat}}}{r^2}\left(\frac{\vec{r}}{r}\right) \tag{2.3}$$

and

$$m_e\ddot{\vec{r}}_e = \frac{Gm_e m_{\text{sat}}}{r^2}\left(\frac{\vec{r}}{r}\right), \tag{2.4}$$

where $m_{\text{sat}}\ddot{\vec{r}}_{\text{sat}}$ and $m_e\ddot{\vec{r}}_e$ are the inertial forces acting on the satellite and the Earth respectively. The position vector of the satellite is equal to the sum of $\vec{r}_e$ and $\vec{r}_{\text{sat}}$, which is conveniently written with respect to the inertial coordinate system allowing for direct differentiation,

$$\ddot{\vec{r}} = \ddot{\vec{r}}_{\text{sat}} + \ddot{\vec{r}}_e, \tag{2.5}$$

to obtain the expression for the acceleration of the satellite. By solving Eqs. 2.3 and 2.4 for acceleration and substituting into Eq. 2.5,

$$\ddot{\vec{r}} = -\frac{G(m_e + m_{\text{sat}})}{r^2}\left(\frac{\vec{r}}{r}\right), \tag{2.6}$$

the second-order, nonlinear, vector differential equation for the two-body problem is found. Assuming the mass of the satellite is small relative to the Earth, Eq. 2.6 is rewritten,

$$\ddot{\vec{r}} = -\frac{\mu}{r^2}\left(\frac{\vec{r}}{r}\right), \tag{2.7}$$

where the gravitational parameter, μ, is the product of G and m_e. When integrated, Eq. 2.7 allows for modeling of satellite position and velocity as a function of time, which will be important to any discussion of satellite constellations, maneuvering, and ultimately reconfiguration. While Eq. 2.7 provides a convenient analytic approximation of the true orbit, its usefulness is limited due to the assumptions made during the derivation.

There is a variety of other forces that perturb the path of a spacecraft away from the idealized assumptions made in Keplerian motion. These include non-gravitational forces like the drag induced by the atmosphere, or solar radiation pressure. Third body effects of the sun and moon can also play a significant role at Geostationary Earth Orbit (GEO) altitude (~35788 km). However, the dominant perturbing force from GEO to LEO arises because the Earth is not a perfect sphere [20].

2.1.3 Modeling Perturbations

The rotation of the Earth causes its mass to bulge at the equator and flatten at the poles, resulting in an oblate spheroid [20]. By excluding this reality, the point mass

assumption used in the derivation of the Keplerian equations can result in significant modeling errors when predicting the position and velocity. Further development of the EOM to include the primary perturbing effect of the oblate spheroid is presented in what follows.

Previously, Eq. 2.7 was developed to allow modeling of the position, velocity, and acceleration of a satellite orbiting a perfect sphere with evenly distributed mass. In order to incorporate the effects of a non-spherical mass distribution, Eq. 2.7 can be recast using a potential function. A potential function is one way to mathematically characterize conservative forces. The proceeding derivation follows Vallado [17].

The gravitational scalar potential function for a spherical body,

$$U_{\text{sphere}} = \frac{\mu}{r}, \tag{2.8}$$

serves as the starting point for the derivation of the EOM. The acceleration from the spherical body's potential is found by taking the gradient of the potential function. Using the gradient of Eq. 2.8 and writing the satellite position vector in the ECI frame using variables *x, y, z* in $\Re^3$ for the individual position components,

$$\nabla U_{\text{sphere}} = -\frac{1}{2}\frac{\mu}{\left(x^2+y^2+z^2\right)^{3/2}}\cdot\left[2x\hat{I}+2y\hat{J}+2z\hat{K}\right] = -\frac{\mu}{r^3}\vec{r}\,, \tag{2.9}$$

the two-body acceleration for the satellite is found, which precisely matches Eq. 2.7 arrived at earlier. Given the existence of a spherical potential, it is also possible to derive an aspherical potential function, *U*, whose gradient subtracted from that of a perfect sphere,

$$\nabla U - \nabla U_{\text{sphere}} = \nabla R = \ddot{\vec{r}} - \nabla U_{\text{sphere}}, \tag{2.10}$$

mathematically captures the perturbing accelerations due to the aspherical body. By expanding this disturbing potential function, *R*, about the point (*x, y,* z) in an infinite series [20], it is possible to isolate mathematical representations of the largest perturbing effects due to the oblate Earth. These zonal harmonics (spheroid of revolution about the Earth's spin axis), and most specifically the part of the expansion associated with the J_2 coefficient, can cause the orbital plane of a satellite to rotate at rates of several degrees per day and cause deviations from the idealized Kepler orbits on the order of 3 km per day. Further, the value of the J_2 coefficient is about 400 times larger than the next-largest coefficient in the series [25]. As such, *R* for this research is limited to the first zonal harmonic term in its series expansion,

$$R = -\frac{\mu}{r}\left(\frac{a_e}{r}\right)^2 J_2 P_2\left[\sin(\phi)\right], \tag{2.11}$$

where P_2 is the Legendre polynomial,

$$P_2\left[\sin(\phi)\right] = \frac{1}{2}\left(3\sin^2\phi - 1\right), \tag{2.12}$$

and ϕ is the argument of latitude (measured from the North Pole). Expanding Eq. 2.10 using Eqs. 2.11 and 2.12, the exact EOM,

$$\ddot{\vec{r}} + \frac{\mu}{r^2}\left(\frac{\vec{r}}{r}\right) + \nabla\left[\frac{\mu}{r}\left(\frac{a_e}{r}\right)^2 J_2\left(\frac{1}{2}\left(3\sin^2\phi - 1\right)\right)\right] = 0, \tag{2.13}$$

are found. Noting the trigonometric relationships,

$$r = \sqrt{x^2 + y^2 + z^2} \tag{2.14}$$

and

$$\sin(\phi) = \frac{z}{r}, \tag{2.15}$$

the final set of second-order, coupled, nonlinear, component-wise differential EOM,

$$\ddot{x} = -\frac{\mu x}{r^3} - \frac{3}{2}\left(\frac{\mu a_e^2 J_2}{r^5}\right)\left(1 - 5\left(\frac{z^2}{r^2}\right)\right)x$$
$$\ddot{y} = -\frac{\mu y}{r^3} - \frac{3}{2}\left(\frac{\mu a_e^2 J_2}{r^5}\right)\left(1 - 5\left(\frac{z^2}{r^2}\right)\right)y \qquad (2.16)$$
$$\ddot{z} = -\frac{\mu z}{r^3} - \frac{3}{2}\left(\frac{\mu a_e^2 J_2}{r^5}\right)\left(3 - 5\left(\frac{z^2}{r^2}\right)\right)z\,,$$

can be integrated numerically to precisely model the first-order secular effects of J_2 on the propagation of satellites. Propagation simply refers to predicting the satellite's state at any future (or past) time enabling one to evaluate its ability to perform its mission(s). In this work, the optimization method can require millions of propagation operations, so in the interest of increasing efficiency an alternative representation of the dynamics shown in Eq. 2.16 is necessary.

2.1.4 Kozai's Analytical Method

The goal of developing alternative EOM to the dynamics shown in Eq. 2.16 is to reduce the computational cost involved in propagating satellite locations. Kozai's method [26] achieves this end by isolating the secular perturbing effects of a central body in several algebraic EOM. Development of this method relies on a more intuitive representation of the orbit known as *classical elements* [23].

To fully specify the position of a satellite at any moment in time, six orbital elements are required. These include four angular measures and two variables that describe the geometry of the ellipse. Following from Fig. 2-2, the inclination, i, measures

the tilt of the orbit plane relative to the equatorial plane. The right ascension of the ascending node, Ω, is the angle between the inertial $\hat{I}$ axis, which is fixed in space in the same direction as the vernal equinox and the point where the orbital path intersects the equatorial plane (with the satellite moving from the Southern to Northern hemisphere). The argument of perigee, ω, measured from the line of nodes to the orbit perigee, locates the point of closest approach. The satellite is located within the orbit by the true anomaly, v, measured from the perigee direction. The last two elements, semimajor axis, a, and eccentricity, e, are half the major axis of the ellipse and a measure of how much the conic section deviates from being circular, respectively. For eccentric orbits, the angular velocity of a spacecraft is continuously varying, which results in an extremely unwieldy expression for v vs. time. A direct consequence of Kepler's second law, first introduced in Section 2.1.1, relates time to position in eccentric orbits more directly.

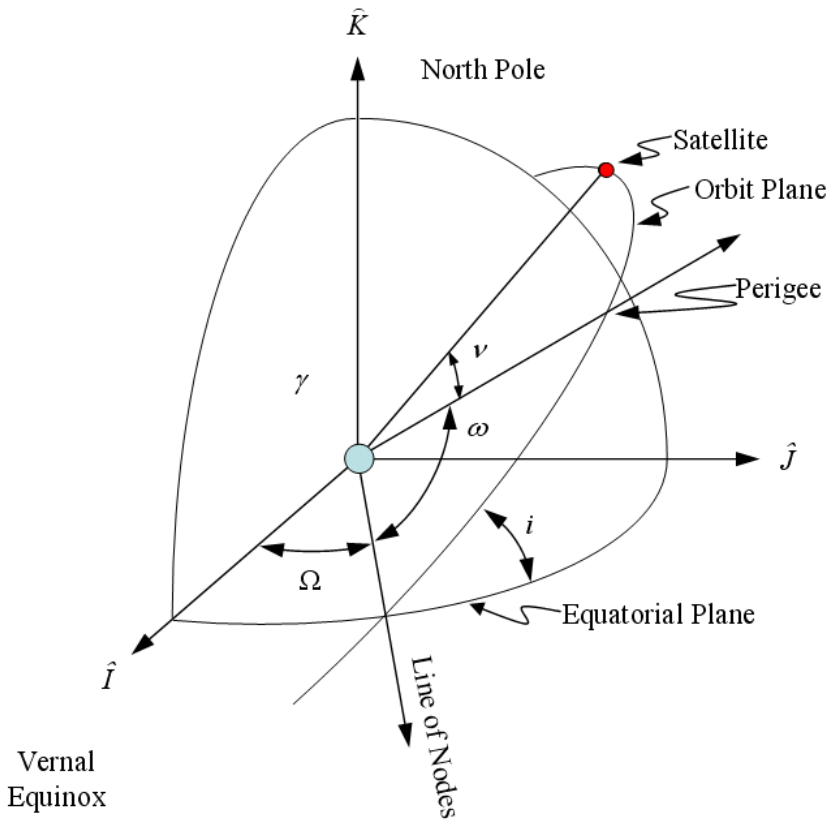


Figure 2-2: Classical orbital elements

Kepler's equation,

$$M = E - e\sin(E) = \sqrt{\frac{\mu}{a^3}}\left(t - t_0\right), \tag{2.17}$$

uses two auxiliary angles known as the mean anomaly, M, and eccentric anomaly, E, (shown in Fig. 2-3) to relate the time-of-flight from epoch, t, and semimajor axis (where t_0 is the epoch time). Note that E is shown in the perifocal coordinate system, $\hat{P}\hat{Q}\hat{W}$,

where $\hat{P}$ points towards perigee, $\hat{W}$ is perpendicular to the orbit plane, and $\hat{Q}$ completes the right-hand system. The mean anomaly physically represents the angular displacement of a fictitious satellite that travels at the mean angular rate as opposed to the true rate, $\dot{\nu}$ [23]. In order to make Eq. 2.17 useful, a variety of geometric manipulations exist where the E is related to ν [24]. The proceeding EOM derivation makes use of two such expressions,

$$\frac{dM}{dE} = \frac{r}{a}$$

and (2.18)

$$\frac{dE}{d\nu} = \frac{r}{a\sqrt{1-e^2}},$$

that relate the differential changes in the auxiliary angles to the classical elements and position magnitude, r.

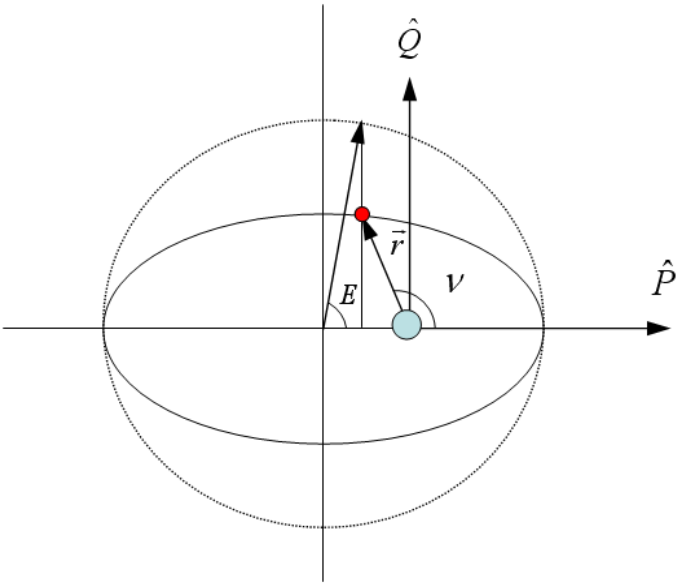


Figure 2-3: Illustration of the eccentric anomaly

Using this set of elements (a, e, i, Ω, ω, M) and the disturbing potential of Eq. 2.11, it is possible to develop Kozai's analytic propagation equations that are much less computationally taxing than numerical integration of Eq. 2.16.

The EOM derivation follows Roy [22] and begins by writing a version of Lagrange's planetary equations,

$$\frac{da}{dt} = \frac{2}{na}\frac{\partial R}{\partial M}$$

$$\frac{de}{dt} = \frac{\left(1-e^2\right)}{na^2 e}\frac{\partial R}{\partial M} - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial R}{\partial \omega}$$

$$\frac{di}{dt} = \frac{\cos(i)}{na^2\sqrt{1-e^2}\sin(i)}\frac{\partial R}{\partial \omega} - \frac{1}{na^2\sqrt{1-e^2}\sin(i)}\frac{\partial R}{\partial \Omega}$$

$$\frac{d\Omega}{dt} = \frac{1}{na^2\sqrt{1-e^2}\sin(i)}\frac{\partial R}{\partial i} \tag{2.19}$$

$$\frac{d\omega}{dt} = -\frac{\cos(i)}{na^2\sqrt{1-e^2}\sin(i)}\frac{\partial R}{\partial i} + \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial R}{\partial e}$$

$$\frac{dM}{dt} = n - \frac{1-e^2}{na^2 e}\frac{\partial R}{\partial e} - \frac{2}{na}\frac{\partial R}{\partial a},$$

where he uses variation-of-parameters [17] to give a set of first-order differential equations that relates the perturbing potential function, orbital elements, and time. Using the relationship between argument of latitude and the classical elements,

$$\sin(\phi) = \sin(i)\sin(\omega + \nu), \tag{2.20}$$

and a trigonometric identity,

$$R = -\frac{3}{2}\frac{\mu}{r}J_2\left(\frac{r_e}{r}\right)^2\left(\frac{\sin^2(i)}{2} - \frac{\sin^2(i)\cos(2(\omega+\nu))}{2} - \frac{1}{3}\right), \tag{2.21}$$

R is rewritten.

Recall the goal in the development of these EOM is to reduce computation cost while capturing the main perturbing variations in the elements. Those terms in R involving short- and long-periodic effects can be segregated and omitted so that only those that contribute to secular variations in the elements remain. Secular effects generally cause linear increases or decreases to a particular element over time and hence are the most significant. The long-period effects oscillate about the secular trend line and are due to trigonometric functions in ω, while short-period variations oscillate about the long-period curves and are due to trigonometric functions of linear combinations of ω and ν. In this research, only the secular effects are modeled because they capture the primary variations in the orbital elements. Using the definition of the satellite's mean motion,

$$n = \sqrt{\frac{\mu}{a^3}}\,, \tag{2.22}$$

the first-order secular disturbing function (due to J_2 only),

$$\widetilde{R} = -\frac{3}{2} n^2 r_e^2 J_2 \left(\frac{a}{r}\right)^3 \left(\frac{\sin^2(i)}{2} - \frac{1}{3}\right), \tag{2.23}$$

is written with the short-periodic term omitted (that dependent on the cosine of $\omega + \nu$).

The inclination does not experience any secular variation since the partials of $\widetilde{R}$ with respect to ω and Ω both equal zero. A periodic term, *r*, still appears in Eq. 2.23. In the interest of capturing the non-periodic variations in the elements due to the perturbations, $\widetilde{R}$, may be averaged over a given orbit revolution. To do this, the part of $\widetilde{R}$ containing *r* is integrated with respect to *M* using Eq. 2.24.

$$\left(\frac{\bar{a}}{\bar{r}}\right)^3 = \frac{1}{2\pi}\int_0^{2\pi}\left(\frac{a}{r}\right)^3 dM \tag{2.24}$$

Using the transformation of variables (the result of multiplying the differential expressions of Eq. 2.18),

$$\frac{d\nu}{dM} = \frac{a^2}{r^2}\left(1-e^2\right)^{\frac{1}{2}}, \tag{2.25}$$

and assuming *e* constant over the interval of integration, *dM* is substituted into the averaging integral,

$$\left(\frac{\bar{a}}{\bar{r}}\right)^3 = \frac{1}{\left(1-e^2\right)^{\frac{3}{2}}}, \tag{2.26}$$

and solved. By substituting Eq. 2.26 into the ratio of $(a/r)^3$ in Eq. 2.23,

$$R_{\text{avg}} = -\frac{3}{2} n^2 r_e^2 J_2 \left(1-e^2\right)^{-\frac{3}{2}} \left(\frac{sin^2 i}{2} - \frac{1}{3} \right), \tag{2.27}$$

the *average* disturbing potential over one revolution, R_{avg}, is found. Accepting R_{avg} as the perturbing function and substituting it into Lagrange's planetary equations and taking the partials with respect to each of the orbital elements, the set of six first-order differential equations,

$$\frac{d\bar{a}}{dt} = 0$$

$$\frac{d\bar{e}}{dt} = 0$$

$$\frac{d\bar{i}}{dt} = 0$$

$$\frac{d\overline{\Omega}}{dt} = -\frac{3}{2} n J_2 \left(\frac{r_e}{\bar{a}(1-\bar{e}^2)} \right)^2 \cos(\bar{i}) \tag{2.28}$$

$$\frac{d\bar{\omega}}{dt} = \frac{3}{4} n J_2 \left(\frac{r_e}{\bar{a}(1-\bar{e}^2)} \right)^2 \left(4 - 5\sin^2(\bar{i})\right)$$

$$\frac{d\bar{M}}{dt} = \bar{n},$$

where $$\bar{n} = n\left[1 + \frac{3}{2} J_2 \left(\frac{r_e}{\bar{a}(1-\bar{e}^2)} \right)^2 \left(1 - \frac{3}{2}\sin^2(\bar{i}) \right) \left(1-\bar{e}^2\right)^{\frac{1}{2}} \right],$$

represent the averaged (indicated by the over-bar) rates of change of the classical elements due to the first-order secular effects of J_2 in the two-body problem. Integrating

each of the first three equations of Eq. 2.28 with respect to time and assuming the satellite state mean values at epoch are $\left(\overline{a}_0, \overline{e}_0, \overline{i}_0, \overline{\Omega}_0, \overline{\omega}_0, \overline{M}_0\right)$, the first three EOM used in this research,

$$\begin{aligned} \overline{a} &= \overline{a}_0 \\ \overline{e} &= \overline{e}_0 \\ \overline{i} &= \overline{i}_0, \end{aligned} \tag{2.29}$$

reveal that in first-order secular theory the averaged orbital elements, $\overline{a}$, $\overline{e}$, and $\overline{i}$ experience no variations. Next, the mean anomaly rate, $\overline{n}$, is integrated with respect to time,

$$\overline{M} = \overline{M}_0 + \overline{n}\left(t - t_0\right)$$

$$\text{where } \overline{n} = n\left[1 + \frac{3}{2} J_2 \left(\frac{r_e}{\overline{a}(1-\overline{e}^2)}\right)^2 \left(1 - \frac{3}{2}\sin^2(\overline{i})\right)\left(1-\overline{e}^2\right)^{\frac{1}{2}}\right], \tag{2.30}$$

indicating that the unperturbed mean motion (n) is increased or decreased by the coefficient of J_2 according to the inclination of the orbit. The rates in node and perigee,

$$\overline{\Omega} = \overline{\Omega}_0 - \left[\frac{3}{2}\overline{n}J_2\left(\frac{r_e}{\overline{a}(1-\overline{e}^2)}\right)^2 \cos(\overline{i})\right]\cdot(t-t_0)$$
$$\overline{\omega} = \overline{\omega}_0 + \left[\frac{3}{4}\overline{n}J_2\left(\frac{r_e}{\overline{a}(1-\overline{e}^2)}\right)^2 \left(4-5\sin^2(\overline{i})\right)\right]\cdot(t-t_0), \tag{2.31}$$

are calculated similarly from the remaining two planetary equations of Eq. 2.28. Note that these average rates of Eqs. 2.30 and 2.31 are a function of $\overline{n}$, which is a direct result of using the relationship,

$$\int ndt = \overline{M} - \overline{M}_0 = \overline{n}(t-t_0), \tag{2.32}$$

during the integration. The EOM, of Eq. 2.29, 2.30 and 2.31 indicate secular motion in three orbital elements $\overline{\Omega}$, $\overline{\omega}$, and $\overline{M}$ that are *linear functions in time* because $\overline{a}$, $\overline{e}$, and $\overline{i}$ are *constant* and equal to their mean value at epoch. In this research, a satellite's future state is determined by solving Eqs. 2.29, 2.30 and 2.31 for some *t* (greater than zero) from epoch. For the propagation duration of the satellites used to demonstrate the general reconfiguration framework, the algebraic EOM provide sufficient representations of reality at a smaller computational cost than numerically integrating Eq. 2.16.

The precise EOM required for any reconfiguration simulation depend on the mission's orbital regime where the propagation will occur and how far into the future the desired state of the satellite is sought. For example, a LEO mission such as those flown by the space shuttle must simulate the perturbing effects of atmospheric drag while a

GEO satellite must be concerned with the gravitational potentials of the Sun and Moon. Over long-duration simulations, solar-radiation pressure, tidal forces, plate tectonic motion, geomagnetism, and others may become important. For further details regarding applied perturbation theory, a comprehensive presentation is given by Chao [21].

2.1.5 Specialized Orbits

The perturbations experienced by orbiting satellites produce opportunities to design *specialized orbits* that have unique characteristics of interest to many missions. One particular type, used in the case study for this research to demonstrate the general framework, is a repeating ground track orbit.

Satellites in repeating ground track orbits return to the same location with respect to the Earth's surface after some period of time such that the path of the spacecraft repeats itself [20]. Satellites that fly in repeating ground track orbits have the primary advantage of viewing the surfaces of the Earth from the same geometry on a periodic basis. This characteristic is important when interpretation of changes in remotely sensed images is required or when the relative geometry between multiple satellites working together must remain periodically constant.

In order for a ground track to repeat, an integral number of orbit periods, p, must occur in an integral number of sidereal days, d (a sidereal day is approximately 86164.10035 seconds). The dimensionless ratio,

$$q = \frac{p}{d}, \tag{2.33}$$

is often [20] used to identify the repeat cycle characteristics for a given orbit. For the orbits detailed later in this research, the satellites traverse exactly 2 orbits per day and repeat their ground track once per day resulting in $q = 2$. To model this, the secular perturbations due to the oblate Earth (Eq. 2.28), must be taken into account.

Recall, the aspherical potential causes the node, argument of perigee, and mean anomaly to drift. To estimate the mean angular motion required for the ground track to repeat,

$$n_{\text{rep}} = q\left(\omega_e - \frac{d\overline{\Omega}}{dt}\right) - \left(\frac{d\overline{\omega}}{dt} + \frac{d\overline{M}}{dt}\right), \tag{2.34}$$

the orbit's q is multiplied by the difference between the Earth's rotation rate, ω_e, and the nodal drift. From this quantity, angular rates of change in argument of perigee and mean anomaly due to oblateness are subtracted resulting in the required mean motion. It is important to note that $d\overline{M}/dt$ from Eq. 2.34 does not include the two-body mean motion, n, of the satellite, rather only the part of the expression that causes a perturbation in the element (the J_2 term from Eq. 2.28). Recalling the definition of mean motion from Eq. 2.22, it is simple to derive Eq. 2.35 for the semimajor axis,

$$a_{\text{rep}} = \sqrt[3]{\frac{\mu}{n_{\text{rep}}^{\ 2}}}, \tag{2.35}$$

of a satellite required to produce the associated n_{rep} for any given orbit. Since the oblate perturbing effects are buried in n_{rep}, solving for a_{rep} is an iterative process. For example, the orbit with $q = 2$, $i = 55^{\circ}$, $e = 0.0$, Eq. 2.35 produces a semimajor axis of approximately 14341 km. This value for a will appear again later when the specific case study is introduced to demonstrate the general reconfiguration framework.

In the previous section, Newton's second law was used to develop the two-body EOM for a satellite of negligible mass orbiting a point mass. The significant departure of the predicted motion resulting from those EOM led to a derivation that approximated the secular perturbing effects of an oblate central body on the classical orbital elements. The result gave a set of algebraic EOM that allow for rapid computation of a satellite's state relative to the alternative of numerically integrating the exact equations. A specialized orbit was introduced where the satellite periodically moves across the sky in the same apparent path as seen by a ground observer. Now that it has been shown how the satellite's motion is modeled and how it is possible to use perturbing effects to produce specialized orbits, the next task is to demonstrate how multiple satellites may be combined together in a *constellation* to best perform their mission.

2.2 Satellite Constellation Design

This research provides a *general* framework for reconfiguration to an *undetermined* configuration B and, as such, it is necessary to outline the basics of satellite constellation design. Multiple satellites distributed in orbit(s) working together to

achieve common objectives form a constellation. The types of satellite constellations are as varied as the missions they serve and, after decades of research, the one truism that emerges is that no absolute rules exist for designing satellite constellations [20]. The reason why there has been no universal constellation design methodology or a one-size-fits-all constellation is because of the generally limitless potential tradeoffs.

For example, consider a LEO communications constellation. At lower altitudes, the lifetime of the mission is potentially shortened due to drag and spacecraft will likely cost less since components do not need to be radiation hardened to the same extent as would be required for a higher altitude. Alternatively, towards the upper range of the LEO regime drag is significantly reduced and fewer spacecraft are required to perform the mission but the harsh space environment becomes a concern. While no single parameter dictates the final design of a constellation, there is one that is generally the principle performance determinant and, as such, has received a great deal of study: coverage.

2.2.1 Coverage

Coverage of a particular location on the Earth, hereafter referred to as a receiver, is defined when a line of sight exists to any satellite in the constellation. Figure 2-4 defines the angular relationships between the satellite, receiver, and Earth-center necessary to develop an algorithm for determining when a satellite covers a receiver.

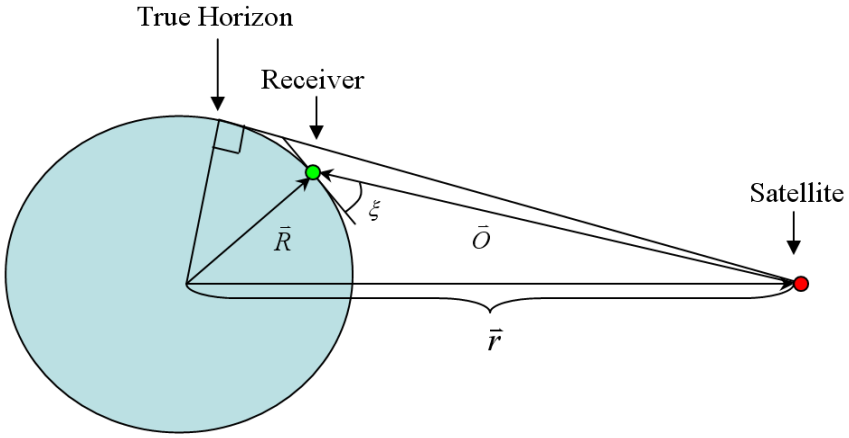


Figure 2-4: Vector relationships between satellite, receiver, and Earth's center, adapted from Wertz [20]

The receiver vector, $\vec{R}$, is measured from the center of the Earth (assumed spherical) to its location on the surface. The observation vector,

$$\vec{O} = \vec{R} - \vec{r}, \tag{2.36}$$

is measured from the satellite to the receiver. The angular measure in Fig. 2-4 is the spacecraft elevation angle (also referred to as the grazing angle in some literature), ξ.

Equations 2.29—2.31 gives the satellite's position in orbital elements at any time from epoch and is easily converted to the position vector in the inertial frame, $\vec{r}$, using the algorithm in Appendix C provided by Vallado [17]. The receiver vector in an inertial reference frame at any time, *t*, is also provided by Vallado [17] and given in Appendix D. A satellite's elevation angle,

$$\xi = \cos^{-1}(\hat{R} \cdot \hat{O}) - 90°, \tag{2.37}$$

is found by subtracting 90º from the inverse cosine of the dot product between the unit receiver, $\hat{R}$, and unit observation, $\hat{O}$, vectors. When ξ is positive and assuming a spherical Earth with no mountains or other obscurations, a line-of-sight from the satellite to the receiver exists and coverage is achieved. In practice, however, it is common to specify a minimum elevation angle, $\xi_{\min}$, at any given receiver below which coverage does not exist. This constraint is incorporated into the oscillatory function,

$$f_{\text{oscil}} = cos^{-1}(\hat{R} \cdot \hat{O}) - 90° - \xi_{\min}, \tag{2.38}$$

by subtracting $\xi_{\min}$ from the expression provided by Eq. 2.37. Consider, for example, a LEO satellite with classical elements shown in Table 2-1 at a t_0 = February 27, 2008 and a receiver located in Los Angeles, California (34.05º North latitude, 118.23º West longitude) with a $\xi_{\min}$ of 5º.

Table 2-1: Classical orbital elements for a single LEO satellite

$\overline{a}_0$ (km)	$\overline{e}_0$	$\overline{i}_0$	$\overline{\Omega}_0$	$\overline{\omega}_0$	$\overline{M}_0$
8000.0	0.0	45°	0°	0°	0°

The EOM defined previously are used to generate receiver and satellite position vectors during a 24-hour period and substituted into Eq. 2.38 to generate the plot of f_{oscil} vs. time in Fig. 2-5. The roots of the function allow for the determination of the times when the satellite *rises* above ξ_{min} and *sets* below ξ_{min}.

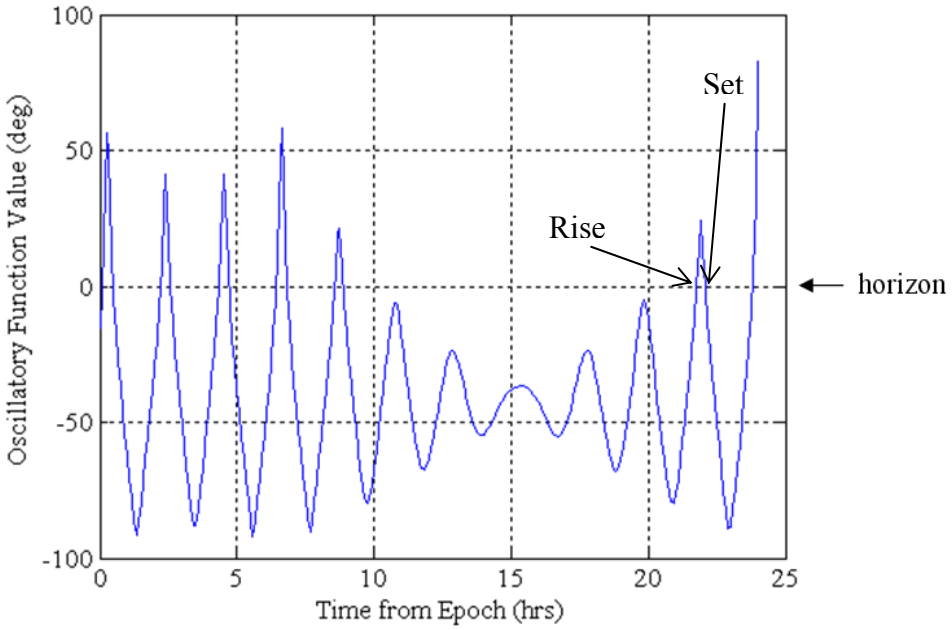


Figure 2-5: Minimum elevation oscillatory access function

The amount of time that passes between the rise and set is known as the access interval. During this time interval, coverage is defined and the satellite has the opportunity (from a purely geometric perspective) to perform its mission. The general approach of writing oscillatory functions to determine access intervals allows for the addition of complex

constraints, which is important for development of the case study undertaken later in this work.

Given an oscillatory access function like that of Eq. 2.38, the rise, t_{rise_h}, and set time t_{set_h}, for access interval *h*, is determined by seeking the time associated with each root of the function. When at least the first derivative of the access function is available, then Newton's method, because of its rapid convergence, is used to determine the zero crossing for each rise and set. However, Newton's method may diverge, and when this is the case, bisection is used to determine the root. While bisection is computationally slow compared to Newton's method, its convergence is guaranteed when the appropriate step size (in time) is chosen. Using these root finding methods (detailed in Appendix E) on Eq. 2.38, the access interval array, shown in Table 2-2 and illustrated in Fig. 2-6, is determined.

Table 2-2: Single-satellite minimum elevation rise/set intervals

h	1	2	3	4	5	6	7
t_{rise} (hrs)	0.0971	2.2135	4.3389	6.4435	8.5545	21.751	23.808
t_{set} (hrs)	0.4759	2.5810	4.7068	6.8221	8.8770	22.077	24.000

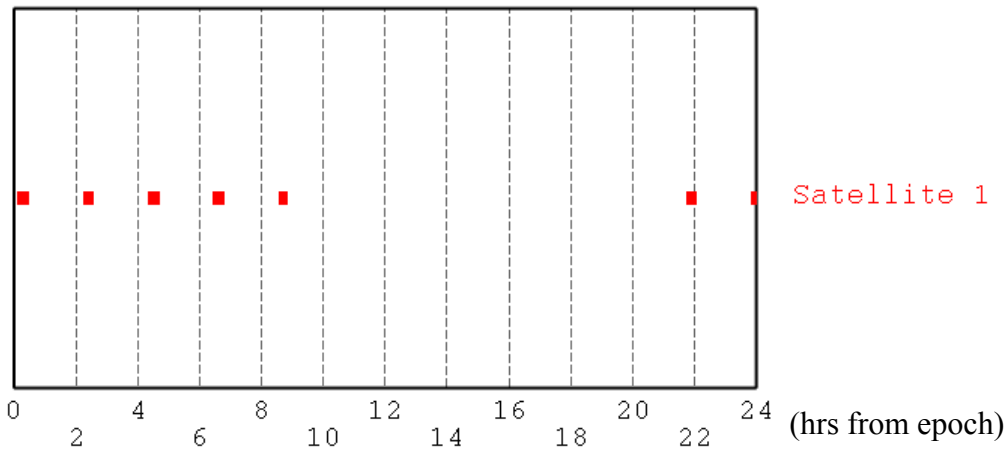


Figure 2-6: Single-satellite minimum elevation access interval plot

These intervals are arrived at by calculating access function values at time steps that are one-tenth of the orbital period of the satellites. This step size is a rule-of-thumb and should be decreased if the oscillatory nature of the function creates sharp, short lived peaks and valleys (otherwise these accesses might be missed). When the sign or slope of the access function changes, the root finding mechanisms are invoked to determine the specific rise or set time. The set of access intervals shown in Table 2-2 serve as the starting point to quantify coverage performance of a satellite to a receiver.

With a constellation, where multiple satellites are working together, the access intervals are combined by the *union* of the individual access arrays. Consider the addition of two satellites to the previous example to form a constellation totaling three spacecraft shown in Table 2-3. The corresponding access interval plots for each satellite and the union of all three are shown in Fig. 2-7.

Table 2-3: Classical orbital elements for a three satellite LEO constellation

Satellite	$\bar{a}_0$ (km)	$\bar{e}_0$	$\bar{i}_0$	$\bar{\Omega}_0$	$\bar{\omega}_0$	$\bar{M}_0$
1	8000.0	0.0	45°	0°	0°	0°
2	8000.0	0.0	46°	0°	0°	15°
3	8000.0	0.0	41°	0°	0°	25°

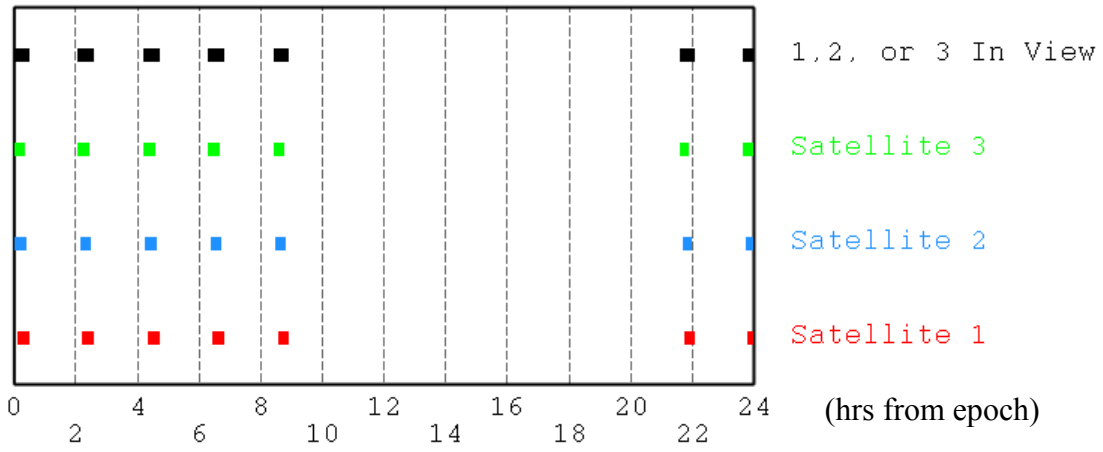


Figure 2-7: Constellation minimum elevation access interval plot

The *constellation* access intervals, from the union of all satellite's accesses, are presented in Table 2-4. This array of rise and set times provides the data needed to quantify coverage performance of the constellation to a single receiver.

Table 2-4: Constellation minimum elevation rise/set intervals

h	1	2	3	4	5	6	7
t_{rise} (hrs)	0.0000	2.0617	4.1849	6.2962	8.4165	21.6076	23.6582
t_{set} (hrs)	0.4758	2.5793	4.7060	6.8228	8.8780	22.0819	24.0000

Some missions, and in particular the case study outlined in Chapter 4, require *F*-fold coverage, which is defined when *F* satellites are simultaneously in view of the same receiver. The access interval array for *F*-fold coverage is determined by performing a Boolean 'and' analysis on the individual satellite access arrays. For the previous three-satellite example, two-fold coverage is obtained when simultaneous access is found for the pairs of satellites 1 and 2, 1 and 3, *or* 2 and 3. The rise/set intervals for two-fold coverage for this example are shown in Table 2-5 and plotted in Fig. 2-8. Note that only the first three hours from epoch are shown in Fig. 2-8 for clarity. The algorithm for calculating the rise and set intervals for multi-fold access is provided in Appendix F.

Table 2-5: Two-fold constellation minimum elevation rise/set intervals

h	1	2	3	4	5	6	7
t_{rise} (hrs)	0.0099	2.1248	4.2507	6.3560	8.4655	21.6683	23.7215
t_{set} (hrs)	0.3884	2.4928	4.6180	6.7348	8.7933	21.988	24.0000

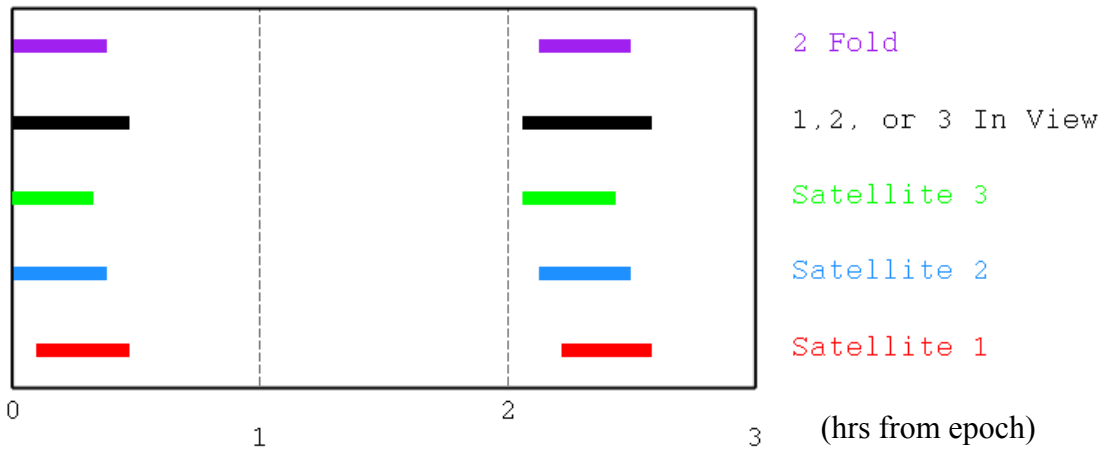


Figure 2-8: Constellation two-fold minimum elevation access interval plot

Now that the building-blocks of constellation visibility to a given receiver have been defined, next it is necessary to assemble them in some manner to quantify, with a single number, coverage performance.

2.2.2 Coverage Figures of Merit

When continuous coverage is achieved, t_{rise_1} is equal to the epoch start time and t_{set_1} is equal to the propagation duration over which the satellites coverage performance is modeled. So the access interval in the continuous coverage case is simply equal in length to the propagation duration. When a constellation experiences a loss, even if it was originally intended to provide continuous coverage, that requirement may no longer be attainable with any possible reconfiguration of the remaining satellites. In this situation, there will be *gaps* in coverage for which the statistics are important. The gaps are simply

the inverse of the accesses, or the durations during the simulation where no access exists. Earth coverage is not an easily characterized statistically distributed parameter [20] (e.g., not Gaussian) and as such it is necessary to accumulate the accesses or gaps into figures of merit, or metrics that allow for the evaluation of the quality of coverage.

There are multitudes of coverage metrics, some of which are tailored for specific missions, while others are applied more generally. Several of the most common metrics are maximum coverage gap, mean coverage gap, percent coverage, Daily Visibility Time (DVT) [20], and mean response time [42]. For the purposes of discussion, the mathematical formulations of these metrics are replaced by short descriptions and the advantages and disadvantages are highlighted. Specific equations used to evaluate coverage performance for the case study in this research are presented in Chapter 4.

The maximum coverage gap, also known in literature as the maximum revisit time [28], is the longest coverage gap encountered for an individual receiver. It is uncommon for a mission to require a constellation to provide access to a single receiver; rather, coverage is sought over single or multiple regions, latitude bands, or the whole Earth, which are all represented by a set of many receivers. Statistics, like the maximum coverage gap are reported as maximum, average, or minimum to the entire receiver set. While this statistic conveys worst-case information, it is rarely used in isolation because of the propensity for a single large outlier gap to misrepresent what might be, on average, a design with much better coverage performance.

The average coverage gap, or average revisit time, is the average length of breaks in coverage for a given receiver on the simulation grid. As with maximum revisit time, the statistics from each point may be combined to report the worst, average, or best case

across all receivers. If the goal is to minimize the maximum revisit time, resources used to revisit any receiver more frequently than any other are wasted [33]. This conflict between the average coverage gap and maximum coverage gap requires a choice to be made between increasing performance in one metric at a cost of a decreased performance in the other. As such, the process of evaluating coverage performance should take into account more than one figure of merit.

The percent coverage for any point in the receiver grid is the percentage of the propagation duration for which access exists. A related metric that captures similar information is the DVT that reports the total duration, during a given solar day, that a constellation has access to a receiver. This metric is particularly useful when a constellation's geometry repeats once per day allowing for a short 24-hour propagation to fully characterize coverage. Both the percentage coverage and DVT do not provide any information about the distribution of gaps but the response time metric does.

The mean response time is the average time from when a random request to observe a point is received until the time when that point is observed [42]. The advantage of using this metric is that it takes into account both the coverage *and* gap statistics and is generally good at evaluating the responsiveness of a given constellation to its receiver set. For the design of discontinuous coverage satellite constellations, geometries are sought that maximize or minimize one or any number of these figures of merit.

2.2.3 Constellation Design Approaches

Constellation design methodologies can be classified into two groups: the first solves continuous coverage problems, the second provides solutions for the discontinuous coverage problem. In previous work, the reconfiguration problem has dealt primarily with communications constellations that have a one-fold continuous global coverage requirement. Towards this end, both Scialom [8] and Siddiqi *et al.* [10] restricted constellation B to geometries described by Lüders [31] and/or Walker [11] continuous coverage design methods.

SOC constellations were first used by Lüders [31] in 1961 and built upon by Adams and Rider [12] in the mid-1980s where they provide a method for calculating the minimum circular polar orbit altitude for continuous global coverage given the number of satellites making up the constellation. The SOC constellation pattern contains SOC_T satellites in each of SOC_P orbit planes such that a continuous 'street' of coverage is formed in each orbit plane. The ideal nodal separation between planes may or may not be symmetric depending on the common inclination of the constellation and at any given time half of the satellites are ascending north while those on the other side of the Earth are traveling south. While SOC patterns are generally a reasonable solution to providing global coverage, the polar patterns result in an intersection of the orbit planes that concentrates resources over regions where few people or facilities exist. Coverage inefficiencies can be reduced by considering patterns that decrease the inclination but still maintain continuous coverage.

One of the major contributions of Walker's work [11] was his systematic approach to producing *symmetric* constellation patterns that assure continuous coverage to any point on the Earth above a specified elevation angle. These *Delta* patterns contain satellites that occupy circular orbits and share common inclinations and altitudes. They are fully defined through the specification of three parameters: the total number of satellites, W_T ; the number of planes, W_P ; and the relative spacing between satellites in adjacent planes, W_F [2]. Determining which patterns provide the best coverage for a given number of satellites is an iterative process usually involving a computer. In 1970, using the Delta pattern recipe, Walker was able to demonstrate that a circular-orbit five-satellite constellation at geosynchronous altitude could provide continuous coverage at or above a minimum elevation angle of 12.3° to every location on Earth. His minimum number of satellite continuous coverage constellation was only improved upon when Draim [30] widened up the design space to include elliptical orbits (he solved the continuous coverage problem with four satellites and received a U.S. patent for the work).

Both Scialom [8] and Siddiqi *et al.* [10] considered how constellation designs might change to support a future higher capacity so their choice to limit the designs for constellation B to those attainable via Walker or SOC patterns is reasonable. However, they did not deal with the discontinuous coverage problem that might arise due to loss in capacity. When this occurs, or if the mission changes to emphasize specific regions, the use of classical continuous coverage design methods are no longer appropriate. In order for the reconfiguration framework to be *general*, the solution to the coverage problem

must be able to provide optimal, or near optimal designs for both the continuous *and* discontinuous scenarios.

It was mentioned previously that the coverage figures-of-merit are non-Gaussian, but more specifically, they can exhibit multimodal, discontinuous, and highly-nonlinear characteristics [13],[28],[32]. Therefore attempting to optimize or find the best discontinuous coverage design for any given mission is not easily solved for using analytical, calculus-based methods. This observation is highlighted by Smith [32] where he compares deterministic (localized) methods that rely heavily on derivatives or gradients (Powell's Method, Steepest Ascent [44]) to non-localized (stochastic) methods (Genetic Algorithm (GA) [43]). Smith found that the deterministic methods converged to incorrect solutions and were sensitive to the initial conditions. On the other hand, he demonstrated that GA was able to avoid pre-convergence to incorrect solutions and also find non-intuitive designs that were "more optimal" than originally thought. Others [13], [33–37] have demonstrated similar success using GA to design both continuous and discontinuous coverage satellite constellations for missions ranging from communications to reconnaissance. However, several recurring drawbacks noted by these authors include the need for extensive parameter experimentation, reliability, and expense both in computational cost and time to convergence.

In 2004 Draim [2] stated that "in [his] opinion, GA hold great promise in designing new constellations." Recently, through the use of parallel processing and advances in the Evolutionary Computing (EC) research community, Ferringer *et al.* [14] and Reed *et al.* [15] have made significant strides towards eliminating the aforementioned drawbacks. These efforts demonstrate that *general* solutions to the

coverage problem must include stochastic methods. Coverage, however, is not the only reconfiguration concern. In what follows, satellite maneuvering is discussed in order to identify the other principlal objective functions of the reconfiguration problem.

2.3 Satellite Maneuvering: Energy and Time

The costs involved in reconfiguring to constellation B form a multi-objective orbital maneuvering problem wherein the goal, as Vallado [17] states, is efficiency. Efficiency can mean using a minimum amount of energy (propellant) on any particular satellite and performing the maneuver in as short a period of time as possible. Two-body analysis of orbital changes typically involves solving Lambert's problem [23] between the position vectors of satellites in constellation A and B.

Many solutions to Lambert's problem have been provided since Lambert's original geometric formulation where he demonstrated how to find the minimum energy orbit between two known endpoints. The transfer is fully specified by the semimajor axis and eccentricity of the transfer ellipse, the initial velocity vector, and the time-of-flight to travel from the position vector in constellation A to its associated vector in B. In Gauss's solution [38] to Lambert's problem, the cost, measured in the energy, ΔV, to perform the transfer in a given amount of time, t_{tran}, is found. Gauss's solution is limited by the type of orbit transfer (elliptical only) and the spread between position vectors (the solution fails when the vectors are far apart). Bate *et al.* [39] detail a universal variable approach that allows for any type of transfer orbit but their method fails on many difficult hyperbolic orbits due the Newtonian iteration used to find the variables. Battin's [40]

solution is not subject to the divergence pitfalls of the previous method and gives the most general procedure valid for any orbit transfer.

While Battin's general solution provides the necessary ΔV, given a time-of-flight, to transfer between *any* two position vectors, several less computationally expensive approximations can be used when mission requirements allow for simplification. When only changes to a, e, and/or ω, are required, then either tangential or non-tangential coplanar maneuvers should be considered. If the mission requires out-of-plane maneuvering, i and Ω can be changed simultaneously or independently. For example, an impulsive burn performed at the nodal crossing will only change i whereas the same burn applied at a specific location in the orbit will only change Ω. In some cases, no change to any of the aforementioned orbital elements is required or permitted leaving rephasing of the mean anomaly, M, within the existing orbits as the only option for reconfiguration.

In what follows, the most common simplifications to Lambert's general solution, classified into coplanar and non-coplanar maneuvers, are presented. The specific maneuvering solutions employed in the framework are a function of the mission under study but should quantify the cost in terms of ΔV and t_{tran} for a given maneuver. The intent of Sections 2.3.1 and 2.3.2 is not to give the mathematical algorithms for calculating ΔV and t_{tran} (given by Vallado [17]) for each maneuver, but rather to provide a roadmap for understanding their advantages and disadvantages. These subsections can be used to select the type of simplified solution (if such a simplification is appropriate) for inclusion in the general framework.

2.3.1 Coplanar Maneuvers

A coplanar maneuver begins and ends with the satellite occupying an orbit of the same plane and has potential to produce changes in a, e, ω, v, and M. During coplanar maneuvering, engine burns are performed that are either tangent or non-tangent to the flight path of the satellite. The type of burn, how many to use, and where in the orbit to perform the burn require trading off ΔV for t_{tran}. Two maneuvers that use tangent burns only are the Hohmann and bi-elliptic transfer.

The Hohmann transfer is a two-impulse minimum change in velocity solution for circular to circular coplanar orbits. Figure 2-9 illustrates the geometry of a Hohmann transfer, where the total energy cost to perform the transfer is the sum of ΔV_1 and ΔV_2. For coplanar transfer between coaxially aligned elliptical obits, maximum efficiency of the burn is achieved by performing the first tangent burn at either the periapsis or the apoapsis of the initial orbit and the second at apoapsis or periapsis of the final orbit for increasing or decreasing the semimajor axis respectively. When the ratio of final to initial semimajor axis for a circular to circular transfer is less than 11.94, the two-maneuver solution provided by Hohmann is superior in terms of a minimum change in velocity solution [17]. In some cases when the ratio is greater than 11.94, a three-maneuver bi-elliptic transfer may use less energy.

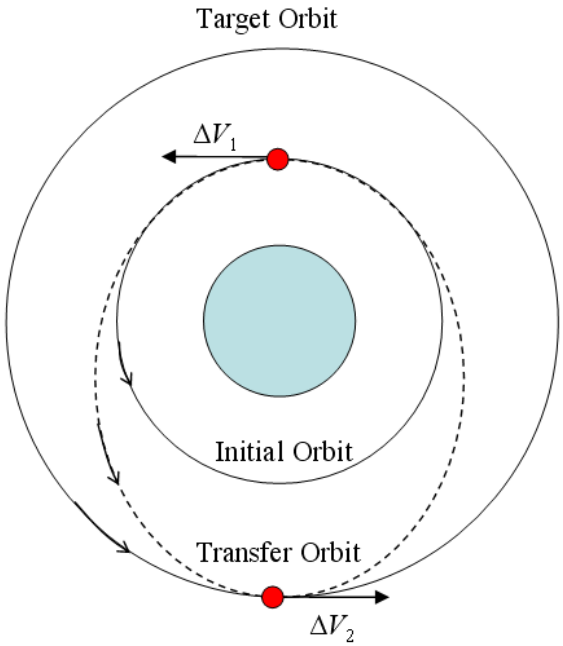


Figure 2-9: Hohmann transfer

Bi-elliptic transfers combine the first burn of a Hohmann transfer and an additional complete two-maneuver Hohmann solution via three impulsive burns. The penalty paid for adopting a bi-elliptic maneuver is a longer t_{tran} relative to a Hohmann maneuver that achieves the same final orbit. A bi-elliptic transfer is shown in Fig 2-10 where ΔV_2 now occurs at the apogee of an extended first transfer ellipse and the third burn, ΔV_3, at the perigee of a second transfer trajectory. If the minimum change in velocity coplanar transfer is sought and the semimajor axis ratio is above 11.94, Escobal [45] describes and provides equations to determine if either a Hohmann or bi-elliptic trajectory should be used.

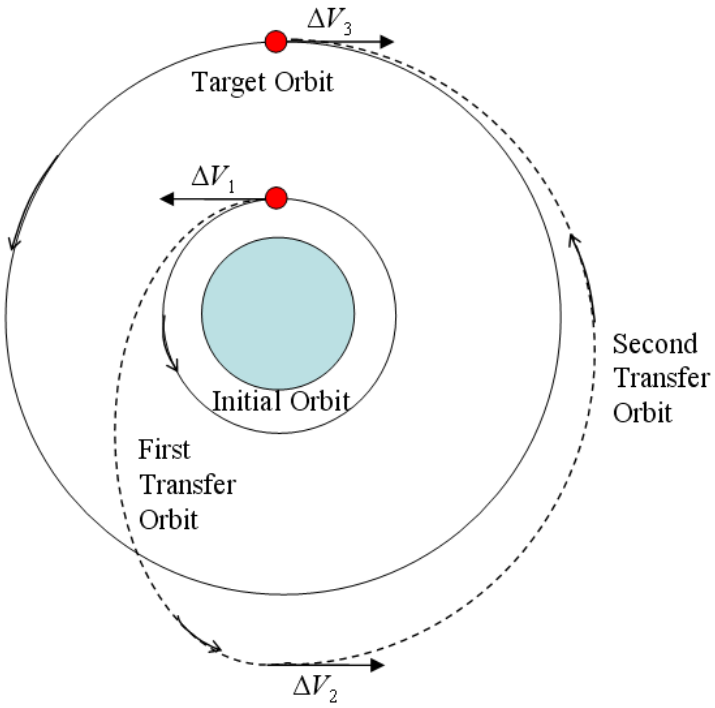


Figure 2-10: Bi-elliptic transfer

There is a third, more general solution that will cost more change in velocity but arrive at the final orbit in a shorter period of time. The one-tangent burn maneuver actually consists of one tangent burn at the apoapsis or periapsis of the transfer orbit and one non-tangent burn at a location in the trajectory that is a function of either a required transfer time or desired propellant expenditure. One-tangent solutions often make use of any of the conic sections for the transfer trajectory. Figure 2-11 provides one such example for an elliptical conic. As t_{tran} approaches zero, the ΔV required goes to infinity so a decision must be made that provides an acceptable compromise between energy expended and time. The one-tangent burn transfer was the type used during the Apollo

program to place crews at the moon in ~3 days instead of the ~5 days required by a Hohmann transfer.[2]

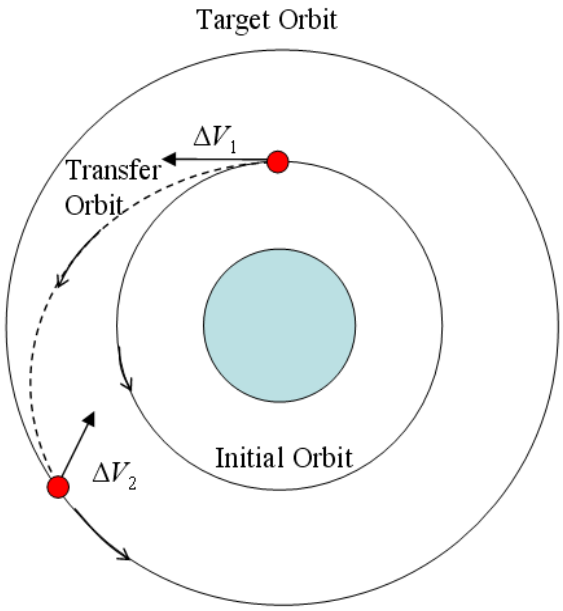


Figure 2-11: One-tangent burn transfer

The final and most general type of transfer consists of two non-tangent burns. Finding the required energy and transfer time in this case requires a solution to Lambert's problem discussed earlier. While the aforementioned maneuvers produce changes in a, e, and ω, it may also be desirable to change the location of the satellite within the same orbit.

Phasing produces changes in M (or ν) and is particularly necessary when considering real-world operations placing satellites into a constellation. Changing the

[2] http://en.wikipedia.org/wiki/Apollo_8, cited October 8, 2008.

position of a satellite within its orbit may be accomplished by either in-plane or out-of-plane maneuvering. Due to its relevance to the case study presented in this research, only circular coplanar phasing will be discussed here.

Coplanar phasing, where the goal is to move from one position to another in the same circular orbit, is an interceptor-target rendezvous problem. The *target* is the desired end point of a satellite in constellation B, which is transferred from the starting point, or the *interceptor*, of constellation A. Two basic scenarios exist wherein the interceptor is either trailing the target, Fig. 2-12a, or leading the target, Fig. 2-12b.

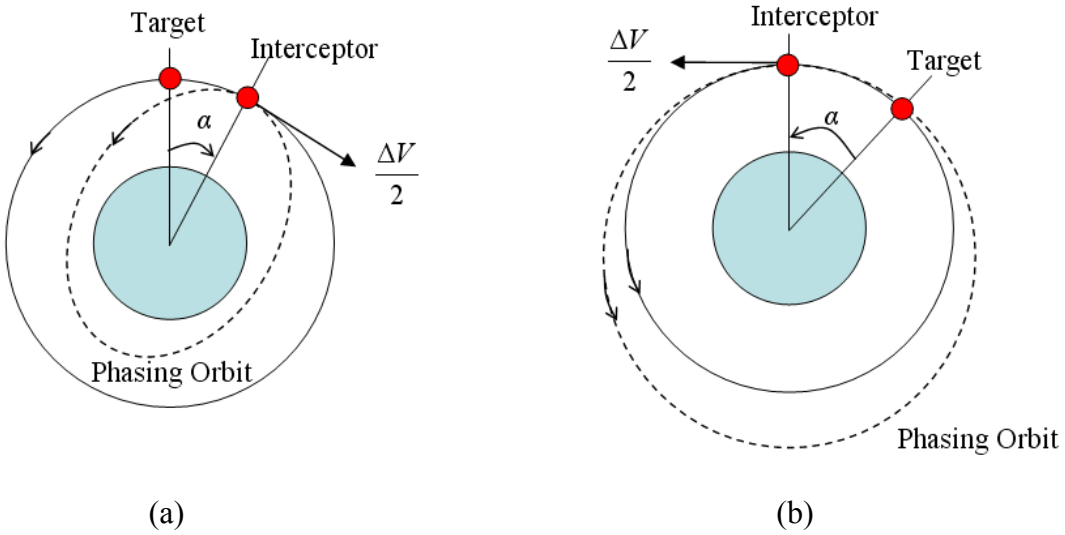


Figure 2-12: Circular coplanar rendezvous (a) interceptor trailing target, (b) interceptor leading target, adapted from Vallado [17]

In each case, the maneuver changes the satellites position by a phase angle, α, relative to the starting position. This is accomplished by applying half of the total maneuver ΔV at the interceptor location to send the satellite into a transfer orbit residing inside or outside

of the nominal circular orbit. The total ΔV required to change M by α is a function of number of integer revolutions, K, which the satellite completes after the initial burn before rendezvousing with the desired interceptor position. In Fig. 2-12a, K can be zero, while with the rendezvous illustrated in Fig 2-12b, K must be greater than or equal to one. With K less than one, the ΔV required to achieve rendezvous can be excessively large. The phase angle is calculated from the change in M from the interceptor to the target and always takes a value between $-360°$ and $360°$ (negative if trailing, Fig 2-12a, and positive if leading, Fig. 2-12b). The satellite must speed up (Fig 2-12a) or slow down (Fig 2-12b) to enter its final position so the remaining half of the total ΔV is applied in the opposite direction at the location of the target after some specific amount of transfer time (a function of K, α, and a of the target orbit) has passed. The specific mathematical relationships to determine energy expenditure and transfer time for circular coplanar phasing will be detailed in Chapter 4. The remaining orbital elements, Ω and i, may be changed by executing a non-coplanar maneuver.

2.3.2 Non-coplanar Maneuvers

Non-coplanar maneuvers are those that apply a ΔV out of the orbit plane and, since they only involve changes in Ω and i, there are three possibilities. The first changes i only through the application of an out-of-plane ΔV at one of the two nodal/equatorial crossings while the second involves an out-of-plane ΔV at a particular point in the orbit which only changes Ω. Such a ΔV applied at any other point in the orbit will result in changes to both Ω and i. Previously, coplanar maneuvers were described with regard to

their energy expenditure *and* time costs to complete transfer. For the proceeding single-impulse maneuvers that change Ω and/or *i*, only energy expenditure is calculated because the change is assumed instantaneous (although in reality all burns are finite).

For both, elliptical and circular orbits, the ΔV cost to change inclination is typically much more expensive than a coplanar transfer. The ΔV required to change inclination by a particular amount is a function of the velocity at the node where the burn is performed, and if the orbit is elliptical, its flight path angle. Analyses for changes to Ω, while also potentially expensive in ΔV, are not as straightforward. A simplified analysis exists [17] for circular orbits that involves applying the impulse at one of the two common points of intersection that occur between the orbit planes before and after rotation through a particular $\Delta\Omega$. With elliptical orbits (with the exception of polar inclinations) multiple burns are required to affect changes in node without altering the other elements like the argument of perigee. Finding energy costs of node changes for elliptical orbits are best handled by solving Lambert's problem. Using spherical trigonometry, it is possible to arrive at simple expressions for combined changes to inclination and node for circular orbits [17]. As before, due to the requirement to correct the argument of perigee for elliptical orbits, it is better to apply one of the solutions to Lambert's problem.

To this point, only impulsive maneuvering has been discussed. Although not relevant to the case study presented in this thesis, another possible transfer becomes important for interplanetary missions or for extended duration maneuvers. These continuous-thrust solutions require the use of optimal-control theory [46] applied to

orbital maneuvering and are beyond the scope of the current discussion. The key point, however, is that, if propellant is limited or if low-thrust engines are available on the satellite undergoing the maneuver, significant energy can be saved by considering a continuous-thrust transfer at the cost of a much longer wait time to achieve the final position.

Any of the aforementioned maneuvering solutions for finding time-of-flight and energy required to perform a maneuver can be incorporated into the general reconfiguration framework. If several simplified maneuvers are needed, they may be performed in sequence or a savings in energy cost could result by combining the sequence into a single transfer. In this case Battin's [40] general solution to Lambert's problem should be used. All of these options were presented so that, when appropriate, the least computationally complex choice for finding the time and energy required for a given maneuver could be selected for inclusion as objective functions in the framework.

To review, three principal concerns that make up the reconfiguration tradeoff-space were identified: coverage performance, energy expended by each satellite during transfer, and time for each satellite to achieve their new position in constellation B. In the next section, several other concerns will be introduced that must be taken into account during any search for an optimal reconfiguration.

2.4 Satellite Maneuvering: Life Expectancy and Risk

When making a decision regarding how to reconfigure a satellite constellation, stakeholders must evaluate the impact on the reduction in life expectancy and risk

undertaken due to maneuvering. A constellation's lifetime can be shortened if one satellite uses most of its remaining propellant relative to the other spacecraft. Previously, Siddiqi [10] attempted to find reconfiguration designs that minimized the sum of the ΔV used by all maneuvered satellites but noted significant operational drawbacks to defining the objective function in this manner. He suggested that future study should consider minimizing the sum of the energy usage variance among all maneuvered satellites.[3] Using this definition, the best reconfigurations would be those where the maneuvered satellites expended similar amounts of energy. In an ideal sense, where all satellites in a given constellation have the same amount of propellant, engine, and dry mass at the time of reconfiguration, the use of energy is appropriate. Operationally, however, it is more likely that, when a constellation experiences a loss, the remaining satellites will have varying amounts of propellant remaining, and potentially different engines and dry masses. In this research, the life expectancy objective function takes both the idealized form from Siddiqi's suggestion, and is modified to account for the more likely operational situation so that the impact on the resulting reconfigurations may be assessed.

Risk for reconfiguration can take many definitions. If only a single satellite is maneuvered to achieve constellation B, the operational requirements for support systems, system downtime, etc. are far less than if every satellite must undergo a transfer. From this observation, a pragmatic definition for risk can be defined as the quantity of satellites

---

[3] Several years before Siddiqi's suggestion, Ahn and Spencer [9] constructed an objective function that minimized the distribution of propellant among the spacecraft for reconfiguration of a formation flying constellation.

that undergo maneuvering. This definition introduces an integer-valued discrete objective function into the framework.

The discrete risk function joins the nonlinear and discontinuous coverage performance metrics and the multimodal maneuvering time and energy functions to create a complex tradeoff landscape where deterministic optimization techniques fail. The constellation design community has demonstrated that EC provides the potential to find or approximate optima for similar landscapes. Building off of their contributions [13], [33–37], the reconfiguration problem is posed as a tradeoff between minimizing risk, energy/propellant usage, and time of transfer, while maximizing life expectancy and coverage performance, wherein the optimum set of solutions is sought via multi-objective EC methods.

Chapter 3

Methodology

In this chapter, Multi-Objective Optimization (MOO) is formally defined providing the background of classical MOO techniques necessary to motivate a subsequent discussion of EC methods. In the previous chapter, a considerable emphasis was placed on driving down function evaluation time via techniques to reduce the complexity of the EOM or with simplified maneuvering solutions. Despite these efforts, the process of optimizing satellite constellations is extremely computationally expensive and steps must be taken to mitigate this issue so that decisions can be made in a reasonable period of time with the resources available. Chapter 3 will also delve into the issues associated with parallelizing and using Multi-Objective Evolutionary Algorithms (MOEA) in practice and detail the specific implementation used in this research.

3.1 Multi-objective Optimization

In contrast to single-objective optimization, multi-objective optimization problems (MOPs) arise when an optimal decision is sought for two or more conflicting objectives. Formally, multi-objective optimization is the process of simultaneously maximizing and/or minimizing k objective functions,

$$\vec{f}(\vec{x}) = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_k(\vec{x}) \end{bmatrix}, \tag{3.1}$$

subject to B inequality,

$$T_j(\vec{x}) \geq 0, j = 1,2,\ldots,B, \tag{3.2}$$

and w equality constraints,

$$V_u(\vec{x}) = 0, u = 1,2,\ldots,w. \tag{3.3}$$

In Eq. 3.1, the decision vector, $\vec{x}$, is a D-dimensional combination of variables that define the design space, Ψ. Each decision vector evaluates to a single point in the k-dimensional $\Re^k$ objective space, Λ. The goal of a multi-objective optimization is to find or approximate the Pareto [47] optimal solutions, $P^* \in \Psi$. Pareto optimal solutions in P^* cannot be improved in one objective without reducing performance in at least one of the remaining objectives. The set of solutions that approximate P^* are said to be non-dominated [48]. When mapped to Λ, P^* defines a Pareto-front (k = 2), Pareto-surface (k = 3), or Pareto-hypervolume (k > 3) that result from the tradeoffs between objectives. It is not just sufficient to find solutions in or near P^* (the first goal of MOO: convergence),

it is also desirable to find solutions along the full extent of the tradeoff space (the second goal of MOO: diversity) thereby providing more choice to a DM. A more formal mathematical introduction to multi-objective optimization and Pareto-optimality can be found in Deb [49].

3.1.1 Classical Methods

The foundations of MOO can be found as early as 1895 with Cantor's work [50] on Set Theory, but it was not until the mid-1960s that *tradeoff* became a common term used by managers and DMs [51]. The first engineering application reported in literature [52] appeared in 1963, but it was not until the mid-1970s that the use of MOO became generalized [53]. Since then, many MOO techniques have been developed and may be classified according to the two stages in which the solution to the multi-objective problem is divided.

The first stage of solving a MOP is a *search* for optimum over several objectives, while the second stage is a *decision* process for what kind of tradeoffs are appropriate and is usually influenced by the problem's stakeholders. Cohon and Marks [51] classify MOO techniques into three categories: 1) *a priori* preference articulation where decisions are made before searching; 2) *a posteriori* preference articulation where decisions are made after searching; 3) progressive preference articulation where search and decisions are integrated in some manner.

A priori methods assume that a decision maker's preference information is introduced prior to the search for optima. Several methods seek the decision vectors that

get as close to an ideal *goal* as possible (Global Criterion Method [54]), minimize the deviation of the objectives from their corresponding goals (Goal Programming [55],[56]), or attempt to under- or over-attain the desired goals sought by the DM (Goal-Attainment Method [57]). Others, like Roa's Lexographic Method [58], rank the objectives in order of preference and proceed by minimizing the most important first, followed by the next in the list until all objectives have been considered. Another method [59] seeks the highest degree of multi-attribute *utility* with respect to all objectives by maximizing over a utility function that mathematically models a DM's preference structure. Many other *a priori* methods exist (see Coello Coello *et al.* [48] for a comprehensive presentation).

Classical *a posteriori* methods do not require preference information from the DM prior to search. One of the most common and easily implemented approach searches for the minimum of a linear combination of the products of weight vector and objective vector components. This Linear Combination of Weights (LCW) [60] approximates the set of non-dominated solutions by parametrically varying the weights and solving each minimization problem. Figure 3-1 illustrates a potential drawback of the LCW method wherein non-convex regions of Λ (between points A and B), for a two-objective minimization problem, are not found due to non-linearity in the mapping between Ψ and Λ. The components of the weight vector, w_1 and w_2 in Fig. 3-1 are those associated with f_1 and f_2.

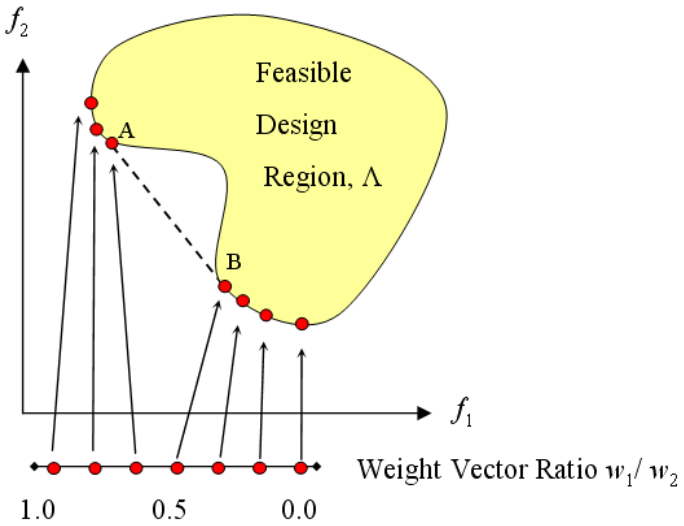


Figure 3-1: Drawback of the LCW method

The ε-Constraint Method [61] alleviates the difficulties faced by the LCW approach in solving problems exhibiting these non-convex regions. Here, the MOP is reformulated where one objective is minimized and the other objectives are incorporated as constraints bound by ε levels. The ε parameter represents an upper bound of the value of the other objectives and does not necessarily mean a small value approaching zero. Through the variation of ε, a non-dominated set can be formed. The final type of classical method integrates search and decision roles.

Progressive preference articulation methods typically assume an initial preference structure provided by a DM, search for non-dominated solutions, and then bring the DM back into the process to iterate on their preferences. This loop is carried out until the DM is satisfied or no further improvement is possible. Several examples include the

Probabilistic Tradeoff Development Method [62], the STEP Method [63], and the Sequential Multi-objective Problem Solving Method [64].

In Chapter 2 it was established that both enumerative and deterministic search techniques are ill suited to provide general solutions to the reconfiguration problem. Evidence for this observation is provided by prior literature [8],[10],[14],[32] and the previous discussion of the characteristics of the principal objectives of the reconfiguration problem. In this section, the classical methods, regardless of their type, dealt with the roles of search and decision making independently and, similar to deterministic optimization methods, many of them imposed limiting requirements on the objective function like linearity or continuity. It is reasonable, then, to suggest that a stochastic method like an Evolutionary Algorithm (EA), is more generally suited to solve the multi-objective reconfiguration problem because they address both search and multi-objective decision making and because their *population-based* exploration allows for any combination of the discontinuous/continuous, discrete, multimodal, and nonlinear objective function characteristics.

3.1.2 Evolutionary Computing Methods

EAs are part of the stochastic class of global search and optimization algorithms (see [48], [65], [66], [71] for examples of other common stochastic paradigms). They mimic Darwinian [101] natural selection by treating possible solutions as individuals in a *population* who compete, mate, and produce offspring. The result of many *generations* of this survival-of-the-fittest reproduction loop is a progression towards an optimal

design, or in a multi-objective sense, a set of $\vec{x}^*$ optimal solutions in P^*. Following the biological analogy, the decision variables are encoded in a *chromosome* using one or many data types in the representation (binary, floating point, symbols, letters). Children are produced from the parents via a *crossover* operator that fulfills the job of perpetuating the characteristics of the fit solutions as the evolution progresses. *Mutation*, which generally occurs at a much lower rate, provides an additional element of randomness during search. These mechanisms work together to simultaneously explore for optimal solutions and exploit knowledge of the fitness landscape. The balance between exploration and exploitation (search and selection) is a classic dilemma studied intensively in Statistical Decision Theory (SDT) [98], [99].

Goldberg [43] compares the EA to the *k*-armed bandit problem [68] in SDT. Take for example, a two-armed bandit, which is nothing more than a slot machine with two arms. Each arm produces an average payout, with one arm paying better than the other. The goal is to maximize return given a particular number of pulls. Initially, nothing is known about which arm will produce, on average, a higher payout, so there is tradeoff between gathering information (exploration) and using that information (exploitation) to accumulate the highest possible reward. The optimal solution for the *k*-armed bandit problem provided by Holland [69] shows *slightly more* than an exponentially increasing number of trials should be given to the observed best arm. If designed properly [43], an EA can give an exponentially increasing number of trials to the high quality, most fit solutions in subsequent generations and, therefore, EAs can be classified as *near-optimal* global search procedures. The first multi-objective

implementation of an EA, the Vector Evaluated Genetic Algorithm (VEGA), was suggested by Schaffer [67] in 1984.

VEGA is a simple extension of a single-objective GA. Illustrated in Fig. 3-2, Schaffer divides the initial population of size *N* into subpopulations equal in size to the ratio of the total population to number of *k* objectives.

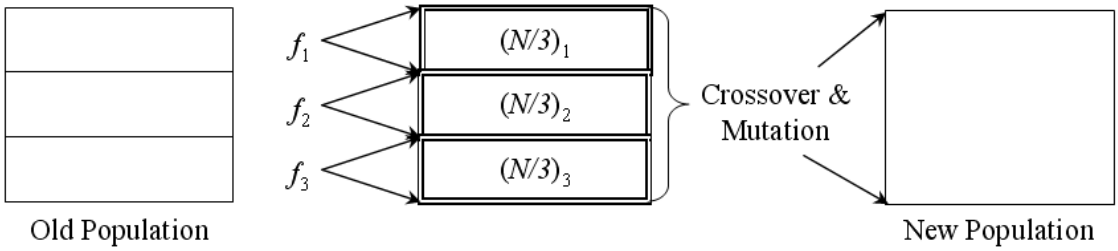


Figure 3-2: VEGA generational flow adapted from Deb [49]

Subpopulations assign fitness according to their particular objective function and crossover is performed across the boundaries. It turns out that performing MOO in this manner produces solutions that perform well in any one objective, thereby meeting the first goal of MOO. By performing crossover between champion solutions of each subpopulation, Schaffer hypothesized that the compromise or tradeoff designs would emerge. Unfortunately, combining the genetic material of two individually fit solutions does not necessarily result in non-dominated compromise designs. It was not until nearly a decade later, in 1993, when an MOEA was implemented that simultaneously met both goals of MOO.

Although first suggested by Goldberg in 1989 [43], Fonseca and Fleming [72] were the first to develop an algorithm, Multi-Objective Genetic Algorithm (MOGA), that explicitly emphasized non-dominated solutions *and* simultaneously maintained their diversity. One year later, Srinivas and Deb [73] published their Non-dominated Sorting Genetic Algorithm (NSGA), which more directly implemented Goldberg's non-dominated sorting concept. That same year, Horn *et al.* [74] proposed the Niched-Pareto Genetic Algorithm (NPGA), which differed from NSGA, MOGA, and VEGA in the selection operator. All three of these mid-1990s algorithms have various advantages and disadvantages [49] and, while they represented progress towards meeting both goals of MOO, they all suffer from the potential for the average fitness of the population to deteriorate over time.

The previous MOEAs did not contain any mechanisms to ensure that the best solutions are retained. The next major advance in the EC community came when researchers implemented *elitism*, which is a mechanism that favors the most fit solutions in a population such that they have the opportunity to be directly carried from generation to generation. Zitler and Thiele [75] proposed an algorithm, Strength-Pareto Evolutionary Algorithm (SPEA), that implements elitism by maintaining an external population of non-dominated solutions that not only preserves the elites but also participates in the genetic operations with the current population. Deb [76] introduced both an explicit elite-preserving strategy *and* a diversity-preserving mechanism in his Non-dominated Sorting Genetic Algorithm Two (NSGA-2).

Due to its relevance as the foundation algorithm extended in subsequent sections for use in this research, the NSGA-2 generational flow, adapted from Deb [49] is provided in Fig. 3-3.

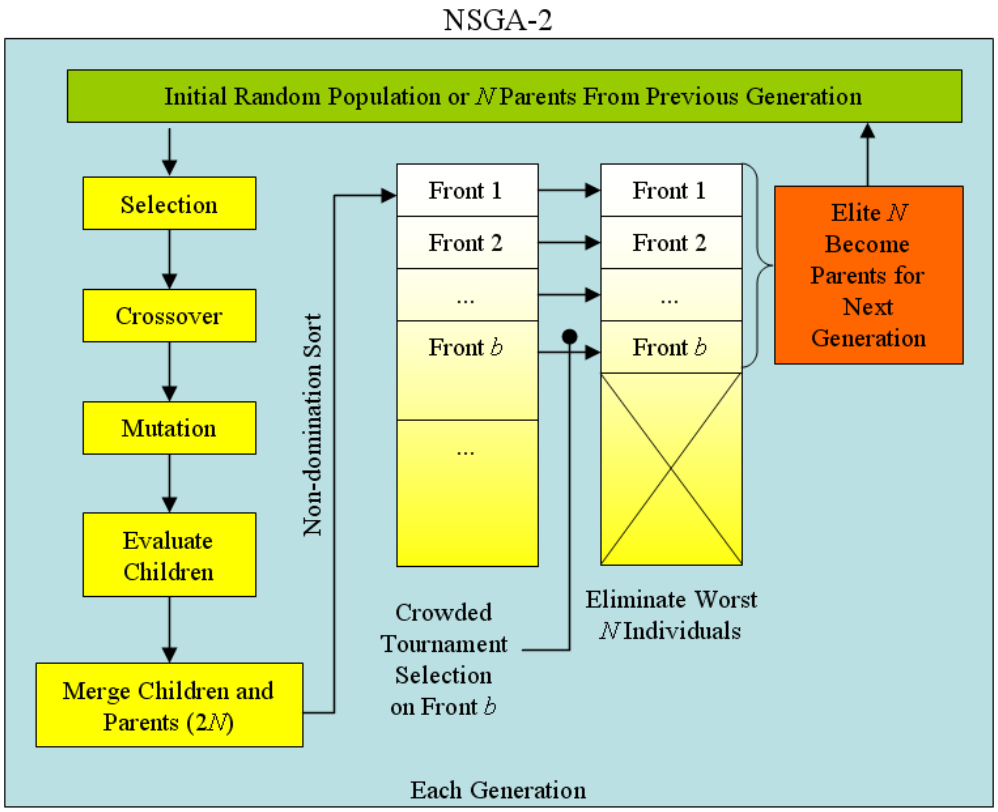


Figure 3-3: NSGA-2 generation flow adapted from Deb [49]

With the NSGA-2, a population of N parents is initially randomly created and combined with a child population generated via crossover [77] and mutation of selected parents. These two subpopulations are merged and non-domination sorting is conducted wherein each solution is sorted into a front according to their domination rank. This

procedure is illustrated in Fig. 3-4 for a two-objective optimization problem, where a merged population of size 2*N* is sorted into four ranks (the coloring of the data markers serve to illustrate the grouping of solutions into domination rank levels). Domination ranking is the first component of an individual's fitness.

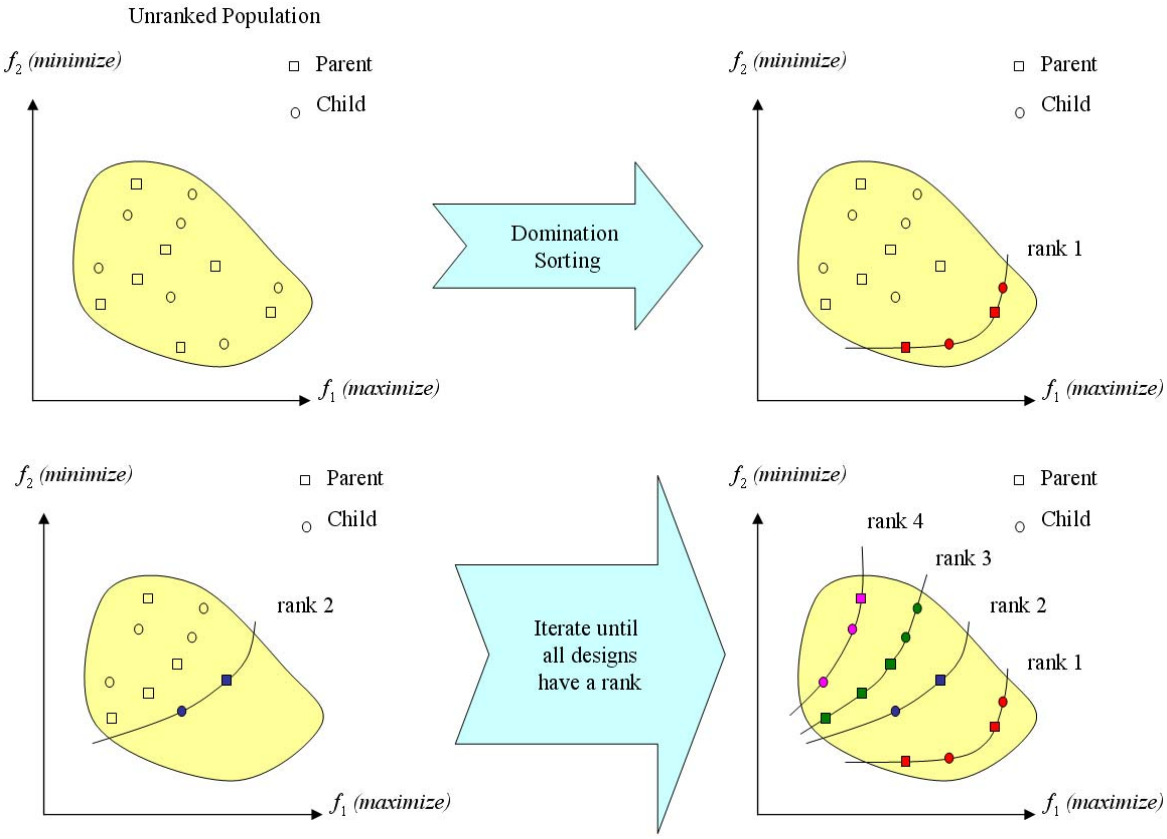


Figure 3-4: Non-domination sorting

The second, crowding distance, provides the diversity-preserving mechanism. Shown in Fig. 3-5 for a two-objective problem, each solution's crowding distance, d_{crowd}, is calculated by averaging the side length of the cuboid formed by using the two nearest

neighbors on either side of the individual as vertices. Deb [49] provides an algorithm that generalizes crowding distance for any number of *k* objectives.

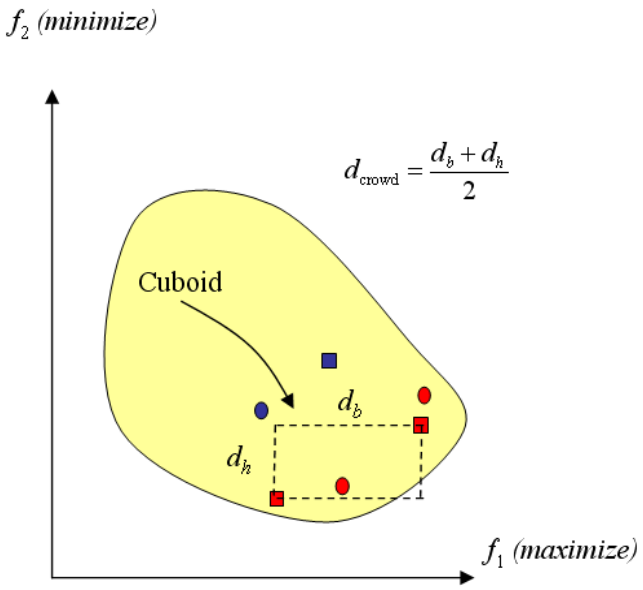


Figure 3-5: Crowding distance

After non-domination sorting, individuals that are members of the most fit (ranks 1, 2, etc.) fronts form a new parent population of size *N*. Since the parent population is fixed at size *N*, any overlapping solutions in front *b* are selected based on their crowding distance. This new parent population begins the cycle anew in another generation and search proceeds until a maximum number of generations has elapsed. As search progresses, all solutions typically become members of the same domination rank and the crowding distance metric becomes important. In this manner, the NSGA-2 can be considered a greedy algorithm that attempts to rapidly approach the Pareto-front early in

the evolution, and enforces the diversity component of fitness only after some convergence is experienced.

While the merging of child and parent populations ensures elitism, because of the fixed population size, the NSGA-2 can loose non-dominated solutions. Further, due to its greediness, the NSGA-2 can pre-converge to a local region in the landscape leading to random seed dependence for successful search. Parameter tuning required for successful search was noted in the constellation design problem domain, first by Smith [32] for a single-objective GA, and then by Ferringer *et al.* [14] for NSGA-2. Additionally, search termination after a number of generations have elapsed is arbitrary and can potentially waste resources if a reasonable solution is arrived at prior to termination. The Epsilon Non-Dominated Sorting Genetic Algorithm Two (ε-NSGA-2) was designed to address these challenges and has compared favorably to the performance of its parent algorithm for a suite of real-world problems [78], [79], [80].

3.1.3 Epsilon Non-dominated Sorting Genetic Algorithm Two (ε-NSGA-2)

The ε-NSGA-2 has three primary innovations (epsilon-dominance archiving [81], auto-adaptive population sizing [80], use of time continuation [82]) that address the previous challenges [14]. Epsilon-dominance sorting, illustrated in Fig. 3-6 for a two-objective minimization problem, allows the user to place a grid over Λ that specifies a particular level of precision, ε_k, for each objective.

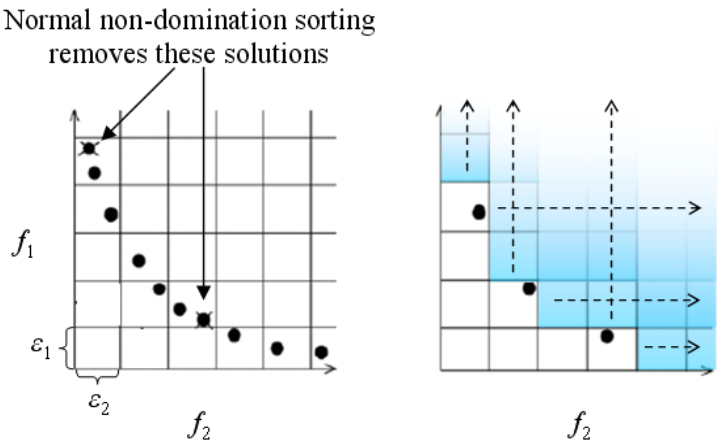


Figure 3-6: Illustration of epsilon non-dominated sorting (a) before and (b) after sorting; adapted from Tang *et al*. [83]

For example, typical non-domination sorting of a given population of designs for a two-objective minimization problem will result in the set shown in Fig. 3-6a. When epsilon-non-domination sorting is performed using the grid formed by ε_1, and ε_2, P^* is reduced to those solutions shown in Fig. 3-6b. In this minimization example, the thinning of points occurs by eliminating every epsilon block that is above and to the right (see Fig. 3-6b) of the boxes that hold the epsilon-non-dominated designs of Fig 3-6a.

The thinning of P^* serves several purposes: convergence towards P^* can occur more rapidly; the decision maker controls the resolution of the final result, which is important since objective values (due to data uncertainties and numerical error) are meaningless beyond a particular level of precision; P^* can only grow to a size limited by the user specified ε_k.

Time continuation is the idea that periodic injection of random solutions will allow the search to continue while maintaining population diversity for as long as computationally feasible or necessary. Auto-adaptive population sizing supports time continuation by growing or shrinking the population to a size proportional to the off-line archive. This innovation eliminates the finite run limitation of NSGA-2 where multiple random seeds might be needed to characterize P^*. Time continuation is particularly important when resource limitations may exclude the possibility of random seed analyses or parameter tuning.

The overview of the ε-NSGA-2, adapted from Kollat and Reed [80], is presented in Fig. 3-7.

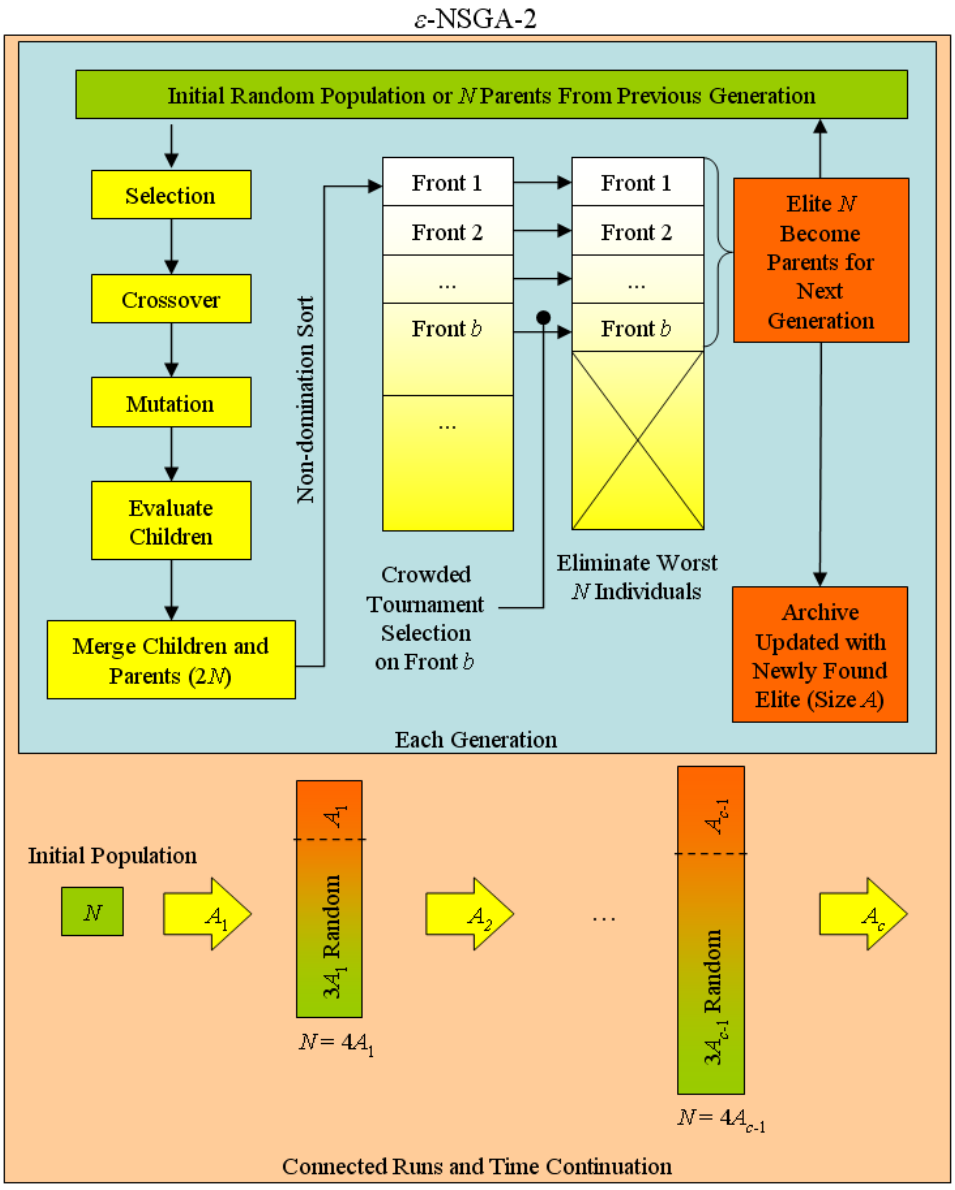


Figure 3-7: ε-NSGA-2 generational flow; final ε-non-dominated population size is A_c at simulation termination; adapted from Kollat and Reed [80]

The ε-NSGA-2 maintains the basic operators (crossover and mutation), and parameter options of the original NSGA-2, but the search progresses differently. With the NSGA-2, the search proceeded from one generation to the next by looping through

the process of non-domination sorting a combined parent and children population of size $2N$; performing crowded tournament selection resulting in a new parent population of size N; then performing crossover and mutation on the selected parents to generate a new child population of size N. The ε-NSGA-2 adds an offline archive that contains A epsilon-Non-Dominated Solutions (ε-NDS) accumulated from each parent population for each generation. The archive is maintained between a series of connected runs, c, shown in Fig. 3-7**,** wherein each run proceeds to a maximum of 250 generations as recommended by Deb *et al*. [76] or until search has failed to produce an improvement of less than 10% in the archive. For each new run, the initial population, N, is generated by taking the archive from the last run (representing 25% of the new population) and injecting $3A_{c\text{-}1}$ randomly generated solutions into the current run. Search proceeds via the connected runs and termination criteria are left to the DM but are generally related to either total elapsed time or lack of improvement in solution quality looking back over one or more runs.

3.2 Parallel Multi-objective Evolutionary Algorithms

The evaluation of the reconfiguration objective vector for this research involves propagating satellites to calculate access-constrained multi-fold coverage over thousands of receiver locations (the specific cases studied will be presented in Chapter 4). On average, one evaluation of this vector requires on the order of one minute per processing core and some scenarios require over 10 million function evaluations (~100 days on a

serial machine) before run termination. This computational demand leads to the consideration of parallel processing.

Van Veldhuizen *et al.* [85] notes that parallelization of an MOEA is motivated by the desire to reduce execution time and/or resource expenditure. In the case of the reconfiguration problem, the motivation is not necessarily to obtain a solution quickly, but rather one of computational intractability. Without parallel computing, arriving at a solution to the reconfiguration problem, as it has been described, simply is not possible in any reasonable amount of time. While there are many ways in which a MOEA may be adapted to use parallel resources, two major paradigms are highlighted in the literature.

These parallel Multi-Objective Evolutionary Algorithm (pMOEA) models are Master-Slave (MS), and Multiple-Population (MP) [86]. The MS paradigm, shown in Fig. 3-8, executes the MOEA on a single *master* core and sends the decision vectors to *slave* cores for evaluation.

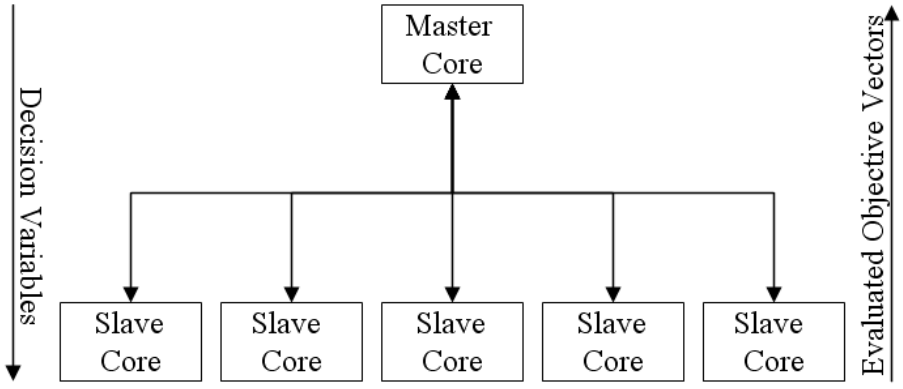


Figure 3-8: MS paradigm

Note, in parallel computing nomenclature a ‘cluster’ is made up of several ‘nodes’ each containing a number of ‘processors’ that may have one or multiple ‘cores’. The MS model searches the fitness landscape in an identical manner to its serial counterpart and requires no new MOEA parameters. The MP paradigm, on the other hand is considerably more complex with regard to parameter choices that affect search.

The MP scheme, shown in Fig. 3-9 for several configurations, is generally modeled after the metaphor that emerges from Darwin’s observations [87] of finch diversity on the Galapagos Archipelago.

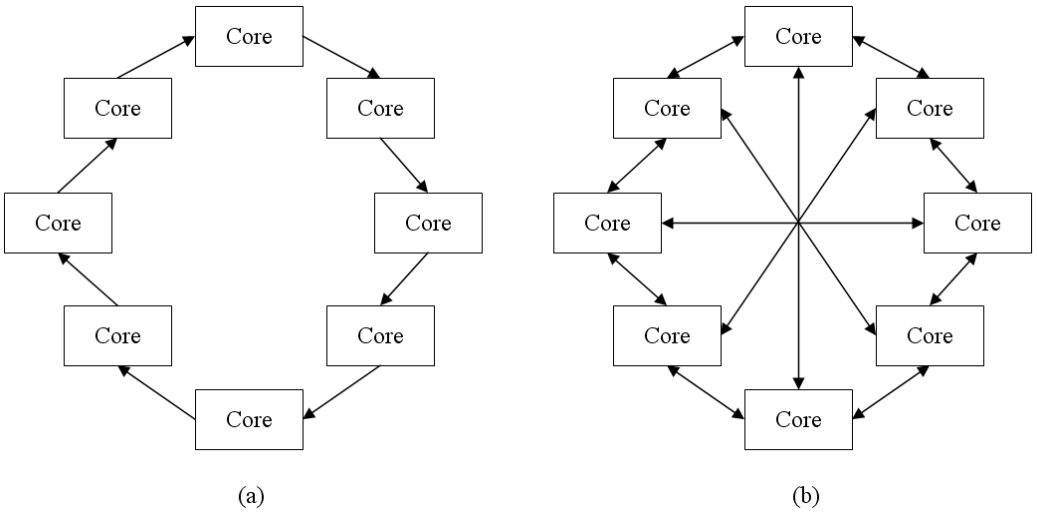


Figure 3-9: MP paradigm for (a) a minimally connected, one-way migration strategy, and (b) for a two-way migration to any of three possible subpopulations

Each core can be thought of as an ‘island’ wherein a MOEA operates in isolation for some period of time before sharing the genetic information of its inhabitants with those of

other islands. The model adds a number of parameters that impact the balance of exploration and exploitation. These include the number of chromosomes to move between subpopulations; the *migration frequency*, or how often genetic information is shared; the number of subpopulations (typically the number of cores); the *connectivity* and *topology* that determine both the number of neighboring islands and how they communicate with each other [85]. Generally, search using the MP paradigm will progress at a slower rate when compared to the MS model for a given period of wall-clock time. The potential advantage, however, is that search progresses differently, which may be more effective for certain problems. To understand why, a slight diversion is provided.

Looking again towards a biological metaphor, the aforementioned MP parameters that control the exploration and exploitation of the search space can be tuned to mimic two contrasting evolutionary theories known as *punctuated equilibrium* [88] and *phyletic gradualism* [89]. Punctuated equilibrium, illustrated in Fig. 3-10a, posits that evolution takes place in rapid bursts separated by long periods in which little change occurs. This theory might be imitated through a low migration rate and a sparsely connected core topology (Fig. 3-9a, for example). Phyletic gradualism, shown in Fig. 3-10b, on the other hand states that evolution occurs uniformly and gradually through the steady transformation of whole lineages.

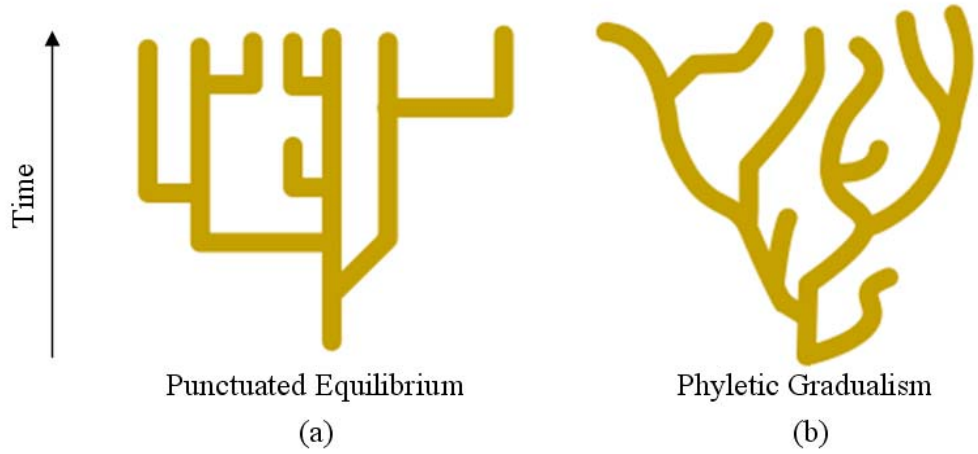


Figure 3-10: Illustration of (a) punctuated equilibrium, and (b) phyletic gradualism

This smooth and continuous view of search progress may correspond to a more fully connected topology (Fig. 3-9b for example) with more frequent migration rates. The MP scheme's multitude of parameter combinations represent exploration and exploitation compromises between the extremes represented by these two theories offering many more potential search paths through the solution space relative to the MS scheme. Accordingly, literature [15] [83] has demonstrated that the MP scheme can dramatically enhance search for extremely challenging problems where the MS model fails. The question then becomes, when little is known *a priori* regarding the complexity of the multi-dimensional reconfiguration problem fitness landscape, which paradigm, MS or MP, is appropriate?

There is no straightforward answer to the previous question. In the early years of this research, lacking any literature that might have provided initial guidance, the approach was simply to implement a MS and MP version of Deb's [76] original NSGA-2

and study their efficiency and effectiveness on a benchmark, two-objective constellation design problem [14]. These two objectives were coverage metrics like those discussed in Chapter 2 and provided initial insight into a subcomponent of the reconfiguration problem. For this initial implementation, Ferringer *et al.* [14] identified and resolved several algorithmic issues associated using a pMOEA on *large* clusters with *heterogeneous* hardware, processing speeds, and architectures. At the same time Tang *et al.* [79] were exploring similar paradigms but with the ε-NSGA-2 on *small homogeneous* clusters to solve multi-objective problems in the water resources domain. Ferringer *et al.* [14] experienced the parameterization challenges, random seed dependence, and finite run limitations of the NSGA-2 noted earlier. The three primary innovations provided by ε-NSGA-2 addressed these issues, but Tang *et al.* [79] had not yet dealt with how their algorithm might scale to large heterogeneous clusters.

Through a joint research effort, Reed *et al*. [15] adapted ε-NSGA-2, merging the contributions of Ferringer *et al*. [14] and Tang *et al.* [79], and used a statistical metric based evaluation framework to compare MS and MP schemes on a benchmark constellation design problem. Several key contributions emerged from their work. The first was a pMOEA, hereafter referred to as the Large-Cluster epsilon-Non-dominated Sorting Genetic Algorithm Two (LC-ε-NSGA-II), which touts all of the benefits of ε-NSGA-II and also scales well with clusters consisting of any heterogeneous hardware and architectures. The second, was that the comparatively simple MS paradigm that exploits time continuation was shown to be statistically superior to the MP scheme. This is due to the fact the benchmark problem was constrained by *time* not the *properties* of the fitness or decision variable landscapes.

Tang *et al.* [83] classifies search failures (i.e. not reaching a user specified level of solution quality) into two categories. The first are problems that have highly deceptive objective and decision space *properties* where no amount of time will allow an algorithm to converge to the global Pareto-set. Tang *et al.* [83] demonstrated this failure with the MS scheme on such a test problem (DZT6 [113], a manufactured test function with many false Pareto-frontiers). The MP scheme, however, did not fail, and was therefore more effective than the MS approach. The second category of search failure is defined when all that is needed for a better solution quality is more computational *time*. The real-world problems of Tang *et al.* [83] in the water resources domain, and the benchmark constellation design problem highlighted in Reed *et al.* [15] both exhibited objective and decision space properties that place them into the *time* constrained category where the MS scheme was shown to be superior.

This insight is applied in the current research where the MS version of LC-ε-NSGA-II is used as the method for solving the heavily time constrained reconfiguration problem. The results presented in Chapter 5 will reinforce the observation that the MS scheme, and its superior parallel scalability, is an appropriate paradigm choice. In the next section, details regarding the LC-ε-NSGA-II adaption [15] of Tang *et al.*'s [79] original parallel version of ε-NSGA-II are provided.

3.3 ε-NSGA-2 Adapted for use with Large Heterogeneous Clusters

Both Tang *et al.*'s [79] original implementation and Reed *et al.*'s [15] adaption of ε-NSGA-II were developed using the Message Passing Interface (MPI) [90] standard.

The development of the MPI specification involved scores of people from organizations in government, industry, and universities, with the first published Version 1.0 of the standard appearing in 1994 and the second major revision MPI-2 in 1997 [91]. A variety of implementations of MPI were transformed into working code, and to-date, well over a dozen implementations of the MPI standard have been developed.[4] The implementation selected for use in this research is Open-MPI, an open source implementation of the MPI-2 standard.[5]

Tang *et al.*'s [79] ε-NSGA-II used the standard form of the MS paradigm wherein the master core sends a single chromosome to each slave for evaluation. As slaves become available, the remaining population members for the current generation are evaluated. The evolutionary operations occur on the master core only after *all* slaves have finished processing. Since the original implementation was designed for use with small (less than 16 cores) homogeneous clusters, the master core also participated in objective function evaluation. There are several scalability issues with this implementation.

Scalability refers to the ability of the MOEA to use available cores to the greatest extent possible as the size of a given cluster increases and its processor speeds and hardware types become more diverse. The two primary scalability issues identified with Tang *et al.*'s [79] implementation include generational synchronization and the master core bottleneck. Generational synchronization refers to the requirement that all function evaluations from slave cores be returned to the master core before evolutionary

[4] http://en.wikipedia.org/wiki/Message_Passing_Interface#Implementations, cited July 17, 2008
[5] http://www.open-mpi.org/, cited July 17, 2008

operations commence. As demonstrated by Ferringer *et al*. [14], when heterogeneity of the cluster and core count increase, the synchronization issue can cause idle time in well over half of the available resources. The master core bottleneck is a result of requiring the master to perform function evaluations as well as coordinate message passing to the slaves. Another scalability issue that may arise when the cluster size increases is a communication bottleneck, particularly when function evaluation time is short. Reed *et al*.'s [15] LC-ε-NSGA-II incorporated the solutions to these scalability issues provided by Ferringer *et al*. [14] into Tang *et al.*'s [79] original MS implementation.

The first, and most obvious adaption is to relieve the master core of its duties to perform function evaluations and only require message passing functions and evolutionary operations. The second, and the key enabler towards eliminating the scalability issue, is the development of *asynchronous evolution* [15], wherein the master core does not wait for the final objective vector evaluations of a given population before proceeding to the next generation. Instead, these individuals are included in a subsequent population. Figure 3-11 presents an overview of the LC-ε-NSGA-II generational flow.

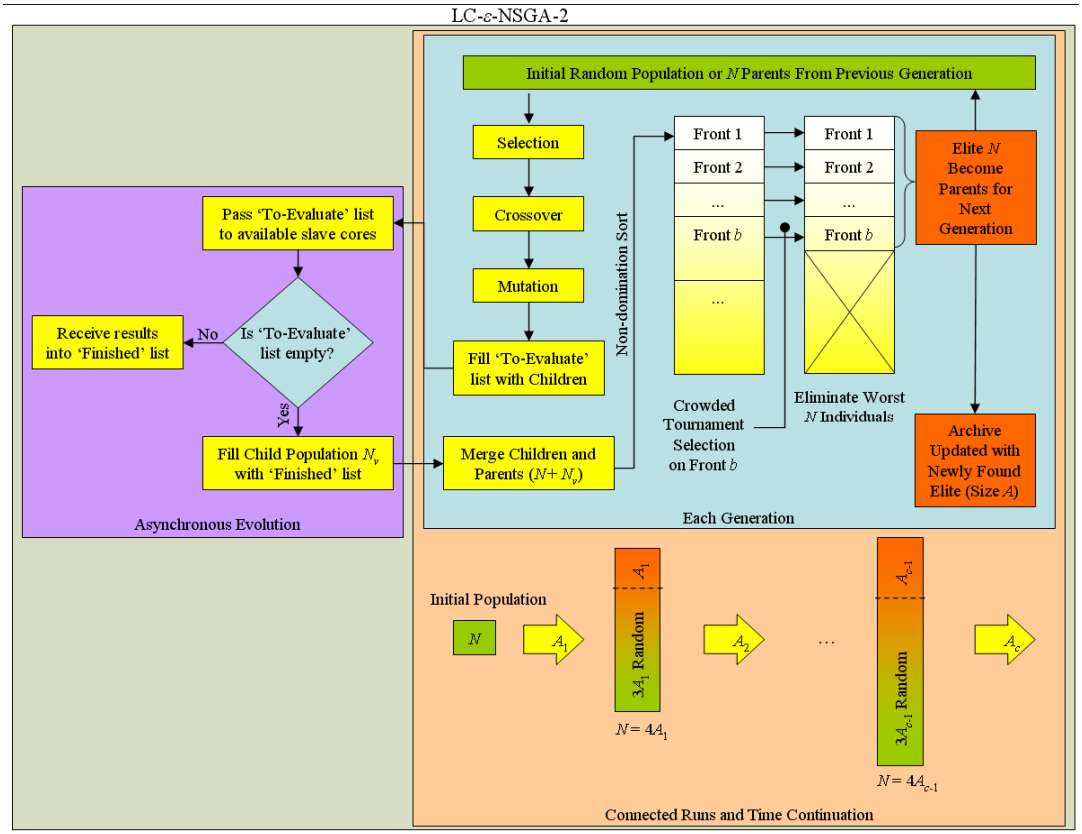


Figure 3-11: LC-ε-NSGA-II generational flow

Asynchronous evolution proceeds by dividing the child population into two lists, one with individuals awaiting evaluation ('To-Evaluate' list) and the other with those whose objective vector has been evaluated ('Finished' list). As soon as the 'To-Evaluate' list has sent its last chromosome to a slave core, those individuals residing in the 'Finished' list form a variable length child population, N_v, that is immediately merged with the parent population N. Even though the child population size varies at the beginning of each run, the population size reaches a steady-state where the number of individuals left behind approximately equals the number of individuals from earlier generations whose

slave cores have finished processing. The net result of asynchronous evolution is near 100% core utilization for clusters containing hundreds of nodes [15]. While the scalability issue is resolved with LC-ε-NSGA-II, another modification was required to cope with potential extend-duration processing time and system failures.

Smith [32] notes when discussing his experience using a GA for constellation design, "as of 1999, computation time issues are a main limitation to the use of this approach in an operational manner." Smith's observation will be supported in Chapter 5, where it is shown that large archive sizes result from the complex high-dimensional fitness landscape of the reconfiguration problem. That, coupled with function evaluation time on the order of minutes per chromosome, even several hundred cores may require weeks to produce a satisfactory solution. In this research, parallel computing resources are shared and limits imposed on the length of time a particular core is used for a given application. To deal with this limitation, a check-pointing mechanism is implemented.

Check-pointing, originally developed [92] to allow for the restarting of a simulation on distributed processors after a fault had occurred, is implemented by periodically saving to disk the population of solutions residing in the archive. When a maximum allowable run time is reached, and if a satisfactory solution has not been obtained, the simulation is resubmitted to the queue of competing jobs and search continues from the last check-point. This mechanism allows for time continuation both with regard to shared resources and any core failure that may occur during the run forcing the simulation to shut down.

Chapter 4

Experimental Design

In this chapter, the experimental design for the general reconfiguration framework is detailed. Principal concerns of performance, cost, and risk are developed into mission specific objective functions to form a six-dimensional tradeoff landscape. A variety of case studies are identified that consider both idealized and operational assumptions. Finally, the specific parameter settings for the LC-ε-NSGA-II with check-pointing are outlined and the parallel computing cluster details are provided.

4.1 Objective Function Construction

The idealized Global Positioning System (GPS) constellation, orbits are pictured in Fig. 4-1, was selected to illustrate the general reconfiguration framework; however, any existing constellation could be used.

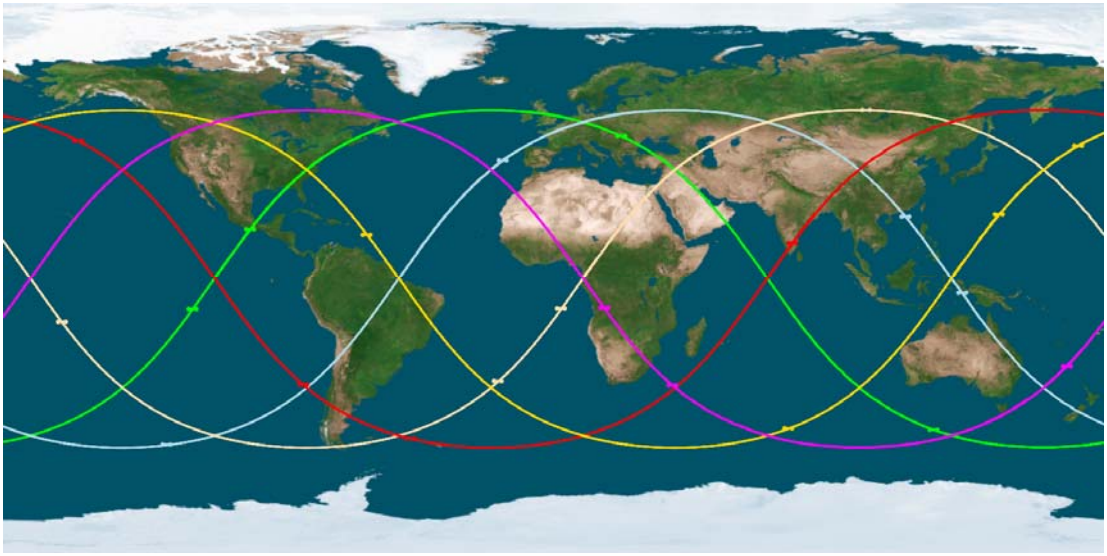


Figure 4-1: Idealized baseline GPS constellation

The idealized baseline constellation consists of 24 satellites in approximately 12-hour circular orbits distributed in six planes inclined at 55° relative to the equatorial plane [93]. The relative phasing between most satellites in adjoining planes is approximately 40° [94]. While the operational constellation includes additional satellites to ensure that maintenance and anomalies will have minimal impact on service, only the nominal 24-satellite constellation is used here. It is not the intent of this research to provide a comprehensive introduction to GPS. However, some background material is introduced to motivate specific objective function definitions.

The principlal goal in developing GPS was to offer users accurate estimates of position, velocity, and time continuously and nearly instantaneously over the globe [95]. By broadcasting ranging signals and navigation data from each satellite, the constellation enables users with a clear view of the sky that have a minimum of four satellites in view

to resolve their position and velocity using receivers. Given the loss of one or more satellites, four-fold global coverage may not be attainable without reconfiguration.

4.1.1 Performance

The performance objective functions are developed with the guidance of the GPS coverage standard performance specification [96]. The GPS constellation must provide greater than 99.9% *global average* and greater than 96.9% *worst-case point* coverage by at-least four satellites above a 5° elevation angle (above the horizon) with a Position Dilution of Precision (PDOP) [95], ρ, less than 6.0. Global average represents a conservative average performance that a user located at any arbitrary location on the Earth can expect to experience. Worst-case point represents a bound on the performance that a user located at the worst possible location on the Earth can expect to experience. PDOP provides a simple characterization of the user satellite geometry, with more favorable geometry producing a lower PDOP. Adopting the notation from Chapter 2, PDOP is calculated,

$$\text{PDOP} = \sqrt{H_{11} + H_{22} + H_{33}}\,,$$

from the geometry matrix,

$$\mathbf{G} = \begin{bmatrix} \left[\dfrac{\vec{r}_1 - \vec{R}_l}{\left\| \vec{O}_{1l} \right\|} \right] \\ \left[\dfrac{\vec{r}_2 - \vec{R}_l}{\left\| \vec{O}_{2l} \right\|} \right] \\ \vdots \\ \left[\dfrac{\vec{r}_S - \vec{R}_l}{\left\| \vec{O}_{sl} \right\|} \right] \end{bmatrix}, \tag{4.1}$$

with

$$\mathbf{H} = \left(\mathbf{G}^T \mathbf{G} \right)^{-1},$$

which contains unit vectors formed by the difference between the satellites position vector and a particular receiver location where $\left\| \vec{O}_{sl} \right\|$ is the range magnitude between satellite *s* and receiver *l*. Note H_{ii} are the diagonal elements of the matrix **H**.

Figure 4-2 illustrates the PDOP calculation during a 24-hour period for four GPS satellites to a receiver located at Vandenberg Telemetry Station (VTS). At t = 0.0 hrs, four satellites (arbitrarily identified by 23A, 24A, 26A, 27A) have at least a 5º elevation angle into the receiver and a PDOP of about 2.0. Some time later, at t = 2.78 hrs, satellite 23A is no longer in view and the PDOP drops to near 3.0. Generally, a better (lower) PDOP will result when more satellites are in view, but it is not always the case. Since PDOP is a measure of the geometric diversity of the satellites in view relative to a receiver, situations can exist where the observation vectors are near or are coplanar,

resulting in a very poor PDOP relative to fewer satellites with a more diverse geometry. Another way to think about PDOP quality is to visualize the volume formed by the receiver and satellites in view. The larger the volume, the smaller the PDOP.

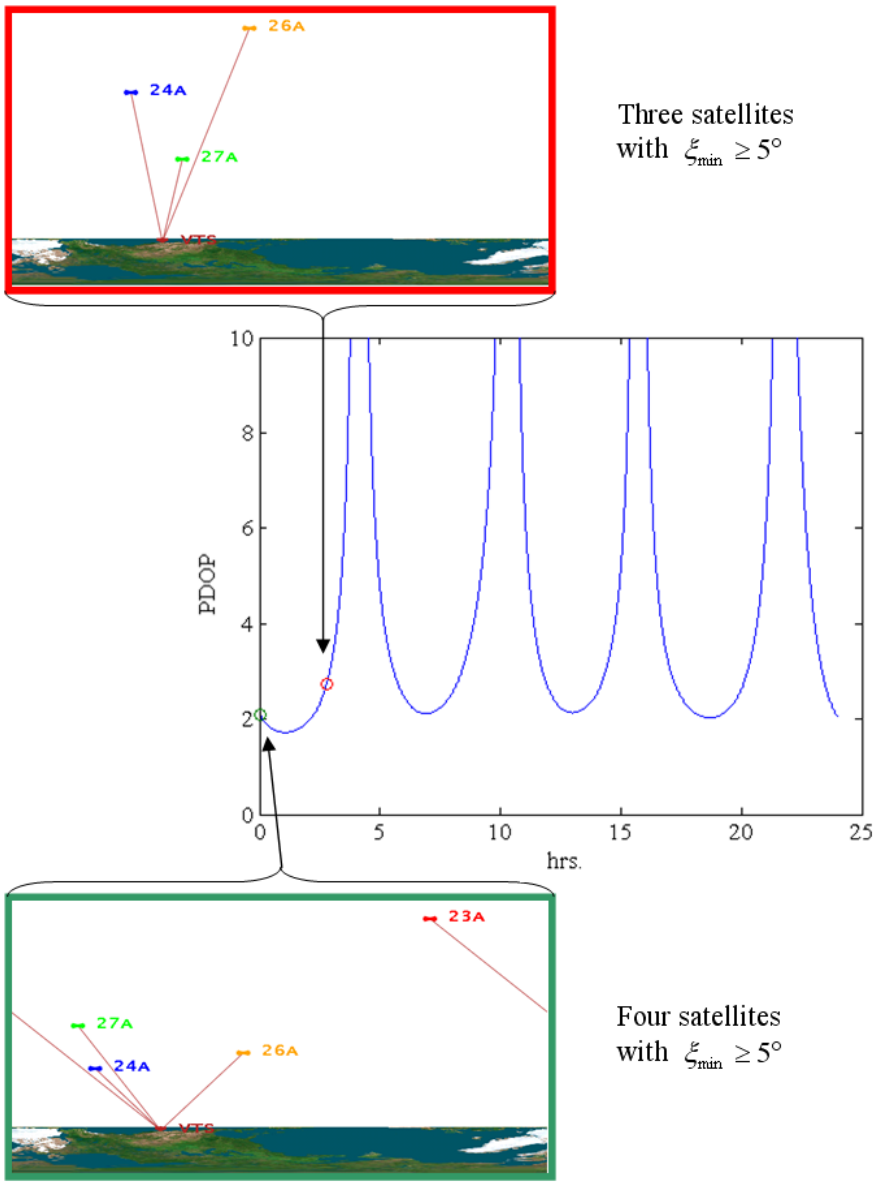


Figure 4-2: PDOP illustration

With GPS, it is sufficient to evaluate coverage by propagating the satellites through a 24-hour period because the approximately 12-hour orbits maintain precisely repeating ground tracks [97]. Coverage performance for any receiver location will be the same whether the satellites are propagated for a day or a year. The next task is to define the receiver grid.

Satellite orbits that have long repeat cycles will experience latitude banding where coverage can be sufficiently evaluated using a small latitude strip of receivers located at a single longitude. For GPS, however, the propagation time is relatively short (24-hours) and 12-hour repeat ground tracks require the *simulation* of a global grid. In an effort to reduce function evaluation time it is assumed that, due to the coverage geometry exhibited by circular orbits, an equal area grid located in the Northern hemisphere only is sufficient to simulate global coverage. The receivers are placed at 0º latitude spaced 3º degrees apart from 0º to 360º East longitude at 0 km altitude. The remaining receivers are placed over the Northern hemisphere such than an equal area grid is formed totaling L = 2339 points. This set of receivers is used instead of an equally spaced grid so that coverage calculations would not be biased towards the polar regions. A grid of this size was chosen as a compromise between taking smaller steps in longitude, (potentially wasting computational resources) versus taking larger steps, where the possibility that considerable coverage gaps might occur between receiver locations is exacerbated.

To capture the coverage standards in the form of objective functions, one or several coverage metrics from those discussed in Chapter 2 must be defined. Towards this end, F-fold Daily Visibility Time, FDVT, is selected, which reports the total time (in

minutes) during a 24-hour period that the constellation has at least F satellites in view of a given receiver. The FDVT coverage function for GPS,

$$\text{FDVT} = \sum_{h=1}^{H} \left(t_{\text{rise}_h} - t_{\text{set}_h} \right) \quad \text{with } F \geq 4, \tag{4.2}$$

is calculated by summing the four- (or greater) fold accesses (refer to Appendix F for the multi-fold access algorithm) to a particular receiver from the one-day propagation. The nominal constellation produces a FDVT of 24 hours for every point on Earth, but a degraded constellation may suffer coverage gaps in various regions. Therefore, the first performance objective function,

$$f_1(\vec{x}) = \max \left[\frac{\sum_{l=1}^{L} \text{FDVT}}{L} \right], \quad \text{such that PDOP} \leq \rho \quad \text{and } \xi_{\min} \geq 5^\circ \quad \text{and } F \geq 4, \tag{4.3}$$

is to maximize the average FDVT to L receivers, subject to a ρ and $\xi_{\min}$ constraint. This objective function serves the goal of attempting to get as close as possible to, or completely restoring, the nominal global average coverage requirement.

The second performance objective function,

$$f_2(\bar{x}) = \max[\min[\text{FDVT}]\ \forall L\ ], \tag{4.4}$$

evaluates a candidate design's ability to meet the worst-case point standard by maximizing the minimum FDVT found across all receivers. Both $f_1(\bar{x})$ and $f_2(\bar{x})$ are, to some extent, in conflict [33] with each other and, as such, it is important to include each in the objective vector so the degree to which they tradeoff can be ascertained.

In general, coverage metrics and their associated access constraints are mission specific. For example, while FDVT constrained by PDOP is important to resolving measurement estimates with GPS, a satellite constellation whose primary goal is to take imagery might use revisit time [13] constrained by minimum image quality to evaluate performance. Recall from Chapter 3 that since MOEAs allow for challenging combinations of objective function characteristics, any mission specific performance metric and constraints can be incorporated into the general framework (limited only by computational demands).

4.1.2 Cost, Idealized Assumptions

The idealized cost objective functions make the assumptions that all satellites have the same amount of propellant remaining, the same dry mass, and the same engine impulse. By allowing part of the reconfiguration cost to be modeled using *energy* instead

of *propellant mass*, the DM is provided with a solution to the question, "Assuming I had unlimited propellant, how should I reconfigure my satellite constellation?" The answer to this question is significant because it sets a performance ceiling for all objectives. In an operational scenario, the lowest ΔV expended by any single satellite may not equate to the least propellant used. That case is handled in the next subsection.

With the idealized assumptions, the cost to achieve the reconfigured constellation B is decomposed into energy usage, time, and mission life expectancy reduction. To formulate the energy and reconfiguration time objective functions, the orbital maneuvers available (refer to Section 2.3 for a comprehensive discussion) must be considered in the context of the mission requirements. The GPS requirement of approximately 12-hour circular orbits resulting in precisely repeating ground tracks rules out changes to altitude and eccentricity. Additionally, changes to inclination or node are very expensive relative to the energy cost of circular coplanar rephasing of the mean anomalies. Since it is desirable to arrive at constellation B with minimal energy usage and time, only maneuvering that results in mean anomaly rephasing is considered.

The first cost objective function,

$$f_3(\vec{x}) = \min\left[\sum_{s=1}^{S}\left(\frac{2\pi K_s + \alpha_s}{\sqrt{\frac{\mu}{a^3}}}\right)\right], \tag{4.5}$$

is to minimize the Total-Time-Of-Flight (TTOF) required to arrive at the reconfigured constellation B. The time-of-flight for each satellite, s, is a function of a, K, and α.

Considering constraints on satellite operators and ground station resources, a conservative definition for the total time for the reconfiguration assumes that only one satellite is moved at a time. The corresponding objective function is to minimize the sum of each time-of-flight for all satellites transferred, *S*.

The next element of cost to be minimized is the largest ΔV required by any single satellite during the reconfiguration,

$$f_4(\vec{x}) = \min\left[\max\left[2\left| \sqrt{\frac{2\mu}{a} - \frac{\mu}{a_{\text{phase}}}} - \sqrt{\frac{\mu}{a}} \right| \right] \forall S \right], \tag{4.6}$$

where the semimajor axis of the phasing orbit,

$$a_{\text{phase}} = \left(\mu \left(\frac{1 + \frac{\alpha_s}{2\pi K_s}}{\sqrt{\frac{\mu}{a^3}}} \right)^2 \right)^{\frac{1}{3}}, \tag{4.7}$$

is needed to determine the ΔV for this objective function. A detailed derivation of ΔV for circular coplanar phasing in identical orbits is provided by Vallado [17].

The final cost objective function,

$$f_5(\vec{x}) = \min\left[\sum_{s=1}^{S}\left(\Delta V_s - \frac{\sum_{s=1}^{S}\Delta V_s}{S}\right)^2\right], \tag{4.8}$$

is formulated to maximize the remaining life of constellation B by minimizing the sum of the variance of the required ΔV among the satellites that undergo maneuvering. This objective function favors designs for constellation B that most evenly distribute the ΔV required by each satellite. In what follows, $f_4(\vec{x})$ and $f_5(\vec{x})$ are reformulated to account for the specific satellite mass/engine properties and propellant remaining at the time of loss.

4.1.3 Cost, Operational Assumptions

For a constellation as large as GPS, where the satellites are launched over decade time scales, design modifications might result in changes to the spacecraft mass and engine performance characteristics. Further, the amount of propellant remaining on each satellite will vary considerably with the oldest satellites near their end of life and other recently launched spacecraft with full tanks. These considerations must be incorporated into $f_4(\vec{x})$ and $f_5(\vec{x})$ to avoid the possibility of an infeasible or unrealizable solution.

Through an application of Newton's second law, where g_0 where the gravitational constant at sea level, the energy an engine will produce,

$$\Delta V = I_{\text{sp}} g_0 \ln\left(\frac{m_i}{m_f}\right), \tag{4.9}$$

is a function of its specific impulse, I_{sp}, and a ratio of the spacecraft mass before, m_i, to that after a burn, m_f [25]. The I_{sp}, or the thrust-per-propellant-weight flow rate, is a measure of propellant quality and typically has units of seconds. Instead of minimizing the maximum ΔV expended by any maneuvered satellite, $f_4(\vec{x})$ is reformulated using Eq. 4.9 and noting that m_u is simply the difference between initial and final mass,

$$f_4(\vec{x}) = \min\left[\max\left[m_i\left(1-\exp\left(\frac{-2\left|\sqrt{\frac{2\mu}{a}-\frac{\mu}{a_{\text{phase}}}}-\sqrt{\frac{\mu}{a}}\right|}{I_{\text{sp}} g_0}\right)\right)\right]\forall S\right], \tag{4.10}$$

to minimize the maximum propellant used, m_u, by any maneuvered satellite.

Previously, the objective function that sought to maximize lifetime was defined to minimize the variance of the energy expended among maneuvered satellites. This function is reformulated,

$$f_5(\vec{x}) = \min\left[\sum_{s=1}^{S}\left(m_{r_s} - \frac{\sum_{s=1}^{S} m_{r_s}}{S}\right)^2\right] \tag{4.11}$$

where $$m_r = m_a - m_i\left(1 - \exp\left(\frac{-2\left|\sqrt{\frac{2\mu}{a} - \frac{\mu}{a_{\text{phase}}}} - \sqrt{\frac{\mu}{a}}\right|}{I_{\text{sp}} g_0}\right)\right),$$

to minimize the variance of the propellant *remaining*, m_r, across all maneuvered satellites. This requires knowledge, or at least an estimate of the propellant mass available, m_a, prior to any burn. The EA generates candidate orbits for satellites in constellation B before the mass required to arrive at the new configuration is determined. As such, it is likely that many potential solutions will be infeasible and, as such, the inequality constraint, $T_1(\vec{x})$,

$$\begin{gathered} \forall S \text{ if } m_r < 0 \\ T_1(\vec{x}) = T_1(\vec{x}) + m_r \end{gathered} \tag{4.12}$$

is enforced. Recall from Section 3.1 that $T_1(\vec{x})$ must be greater than or equal to zero. For all satellites maneuvered that have expended more propellent than was available, m_r will be negative. Equation 4.12 accumulates these infeasible maneuvers such that the more negative $T_1(\vec{x})$ becomes, the greater the constraint violation.

The final task to fully define the operational assumptions is to specify I_{sp}, m_a, and m_i of each satellite in the GPS constellation at the time of loss. The specific values for all of these parameters are not published and, as such, they are estimated with guidance taken from Wertz and Larson [102]. A typical I_{sp} for a maneuvering monopropellant hydrazine thruster is between 220 and 240 sec. It is assumed that, midway through the construction of the nominal 24-satellite GPS constellation, an improved thruster was designed and flown on half of the fleet. Therefore, it is assumed that half of the satellites contain thrusters capable of 220 sec I_{sp} with the other half hosting the more capable 240 sec thruster. The mass of propellant available on each satellite is a function of the duration on orbit, but may also be affected by an unplanned maneuver early in life due to a missed target orbit during launch. As such, m_a is pulled randomly from a uniform distribution with amplitude equal to the maximum beginning-of-life propellant mass budgeted for maneuvering set arbitrarily to 25 kg. In the figures that follow in Chapter 5 it is convenient to define propellant depleted before the maneuver, m_d, as the difference between the beginning-of-life propellant and m_a. Finally, m_i – which is the sum of the dry mass of the satellite, the propellant available for maneuvering, and any other propellant remaining onboard for mission events such as de-orbiting – is given one of two values, 1106 kg or 1136 kg. The assumed values assigned to these variables for each satellite are summarized in Table 4-1. In the event of an actual loss, this table would be populated with precise values or accurate estimates of state of the constellation prior to reconfiguration.

Table 4-1: Assumed operational thruster and mass characteristics prior to reconfiguration

Satellite, s	I_{sp} (sec)	m_a (kg)	m_i (kg)
1	220	17.0	1106
2	220	18.9	1106
3	240	18.6	1136
4	240	9.8	1106
5	240	16.4	1136
6	240	4.3	1136
7	220	17.7	1106
8	240	0.8	1106
9	220	6.9	1136
10	220	1.2	1106
11	240	2.4	1136
12	240	20.6	1106
13	220	17.4	1136
14	220	7.9	1136
15	220	23.8	1136
16	240	0.9	1106
17	220	11.0	1106
18	240	9.5	1136
19	220	19.1	1106
20	220	19.9	1106
21	220	4.7	1106
22	240	12.2	1136
23	240	11.1	1136
24	240	16.2	1136

4.1.4 Risk

The quantification of risk can be highly subjective and constructed from one or many models. One possible pragmatic definition from the perspective of maneuvering can be the integer number of satellites that are actually maneuvered. Each satellite that is transferred requires planning and execution wherein mistakes may be made or

malfunctions may occur. It is desirable, therefore, to limit the number of satellites that undergo maneuvering to achieve the reconfigured constellation B. The following objective function

$$f_6(\vec{x}) = \min[S] \tag{4.13}$$

encapsulates this definition of risk. Moving every satellite in constellation A to achieve B poses the highest risk, while moving only one satellite to arrive at constellation B poses the least risk. Note this sixth objective function takes discrete values introducing yet another challenging objective function characteristic into the existing multi-dimensional landscape.

For reference, each of the numbered objective functions in all figures and tables that follow are identified by the abbreviations shown in Table 4-2.

Table 4-2: Summary of objective function descriptions and abbreviations

Function Number	Optimization Direction	Name	Abbreviation	Units
$f_1(\vec{x})$	Maximize	Four-fold Average Daily Visibility Time	A-DVT	min
$f_2(\vec{x})$	Maximize	Four-fold Worst-Case-Point Daily Visibility Time	WCP-DVT	min
$f_3(\vec{x})$	Minimize	Total-Time-Of-Flight	TTOF	days
$f_4(\vec{x})$	Minimize	Maximum ΔV or m_u required by any maneuvered satellite	Worst ΔV or Worst m_u	m/s kg
$f_5(\vec{x})$	Minimize	Sum of the ΔV or m_r variance of those satellites maneuvered	ΔV Var or m_r Var	m/s kg
$f_6(\vec{x})$	Minimize	Satellites maneuvered	Man	-

4.2 Reconfiguration Case Studies

Two scenarios are developed to illustrate the framework. The multi-objective reconfiguration problem is solved for the loss of a single satellite, Case-1, and for the loss of an entire plane, Case-2. While Case-1 may be brought on by on-orbit malfunctions or natural causes, the second case might present itself if debris causes the breakup of a single satellite resulting in the cascading pollution of a given plane and subsequent collisions. Two optimizations will be performed for each case, where satellite accesses are constrained to a PDOP of 6.0 (per the specification) and 2.0 resulting in a total of four Pareto-hypervolumes. While a PDOP of 6.0 is the standard published in the GPS specification [96], there are both civil [103] and military [104] applications that benefit from a better PDOP, so it is reasonable to contrast changes in the hypervolumes when the

access constraint is tightened. All four of these scenarios are studied for the situation when *K* is common to all satellites or independent. In practice, each satellite will have an independent *K*, but this situation doubles the number of decision variables when compared to a common *K*. By considering both decision vectors, observations can be made regarding the impact of decision vector length on search progress. All eight of these scenarios are considered using both the *ideal* (unlimited energy, identified in the case name by 'ideal') and *operational* (identified in the case name by 'op') assumptions detailed in Sections 4.1.2 and 4.1.3 respectively. The test matrix shown in Table 4-3 depicts the sixteen cases and their associated decision variables for each reconfiguration scenario. For a real scenario, the satellite or plane that is lost is not known beforehand, so the selection of these satellites in Table 4-3 is arbitrary.

Table 4-3: Decision vectors for each reconfiguration scenario

Case-	ρ	Decision Vector	Decision Variables (D)	Index of Satellites Lost (s)
1a-ideal	6	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{23} & K \end{bmatrix}^T$	24	24
1b-ideal	2	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{23} & K \end{bmatrix}^T$	24	24
1c-ideal	6	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{23} & K_1 & K_2 & \dots & K_{23} \end{bmatrix}^T$	46	24
1d-ideal	2	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{23} & K_1 & K_2 & \dots & K_{23} \end{bmatrix}^T$	46	24
2a-ideal	6	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{20} & K \end{bmatrix}^T$	21	21,22,23,24
2b-ideal	2	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{20} & K \end{bmatrix}^T$	21	21,22,23,24
2c-ideal	6	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{20} & K_1 & K_2 & \dots & K_{20} \end{bmatrix}^T$	40	21,22,23,24
2d-ideal	2	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{20} & K_1 & K_2 & \dots & K_{20} \end{bmatrix}^T$	40	21,22,23,24
1a-op	6	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{23} & K \end{bmatrix}^T$	24	24
1b-op	2	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{23} & K \end{bmatrix}^T$	24	24
1c-op	6	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{23} & K_1 & K_2 & \dots & K_{23} \end{bmatrix}^T$	46	24
1d-op	2	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{23} & K_1 & K_2 & \dots & K_{23} \end{bmatrix}^T$	46	24
2a-op	6	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{20} & K \end{bmatrix}^T$	21	21,22,23,24
2b-op	2	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{20} & K \end{bmatrix}^T$	21	21,22,23,24
2c-op	6	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{20} & K_1 & K_2 & \dots & K_{20} \end{bmatrix}^T$	40	21,22,23,24
2d-op	2	$\vec{x} = \begin{bmatrix} M_1 & M_2 & \dots & M_{20} & K_1 & K_2 & \dots & K_{20} \end{bmatrix}^T$	40	21,22,23,24

The decision vectors are made up of the mean anomaly, M, for each satellite and either a common number of integer phasing orbits (Case-1a, -1b, -2a, -2b) or independent phasing orbits (Case-1c, -1d, -2c, -2d). The decision variable ranges are,

$$0^\circ \le M < 360^\circ$$

and

$$2 \le K < 50\,, \tag{4.14}$$

where the choice of limits for K were determined by noting that below 2 revolutions, the ΔV required for rephasing any single satellite becomes prohibitively large while a K above 50 results in an excessive transfer time. For each satellite that undergoes a maneuver, α_s is calculated,

$$\alpha_s = M_{sB} - M_{sA}, \tag{4.15}$$

by taking the difference between M in constellation A, M_{sA}, and the target M in constellation B, M_{sB}.

Due to the GPS constellation's robustness, the constellations with 23 satellites exhibited in Case-1a and Case-1c are able to achieve four-fold 24-hour per day coverage at every receiver location at a PDOP of 6 or less without undergoing any reconfiguration. These cases were designed to test the pMOEA's ability to find a known result: the expert decision is the single Pareto-optimal design representing no action on the part of the DM. For these cases, what the LC-ε-NSGA-2 has to discover is the original 23-satellite configuration at the time of loss of the single satellite. Finding this solution, however, is not trivial because the entire design space must be considered just as it would be with any of the other cases. For the remaining cases, either the PDOP constraint or the loss of more than one satellite requires reconfiguration and the tradeoff between the six objectives is sought.

4.3 LC-ε-NSGA-2 Parameter Settings

Just as with its predecessor [76], both a crossover and mutation probability that provide a balance between search and selection pressure must be specified within the LC-ε-NSGA-2. As recommended by Deb [49] and Kollat and Reed [80], these parameters are detailed in Table 4-4 along with the initial population size.

Table 4-4: Parameter summary

Parameter	Setting
Initial population size	12
Simulated binary crossover probability	0.9
Polynomial mutation probability	1/D
Epsilon vector	[1 min, 1 min, 1 hr, 5 m/s (kg), 1 m/s (kg), 1]

Recall epsilon settings are required that define a grid over the k-dimensional objective space and serve the purpose of thinning the non-dominated set, potentially decreasing time to convergence.

The process of defining an epsilon vector can proceed in the following manner. For example, consider the TTOF objective function, $f_3(\vec{x})$. If a DM would like to retain designs that are at least different by one hour, then an epsilon setting of one hour should be used. Keep in mind, however, depending on where the non-dominated designs fall in the hyper-grid, it is possible (and likely) that solutions will be almost adjacent to each other (meaning there will be much less than an hour difference between them in the case of this example). If one minute were used, many more candidate solutions would be

generated slowing the search algorithm and burdening the DM with an exorbitant number of choices. If instead one week is used, the final set of candidate solutions may be too sparse, resulting in too few choices for the DM. The epsilon vector shown in Table 4-4 is used for the majority of this research; however, an experiment is presented in Chapter 5 that studies the impact of epsilon value selection on the resulting hypervolumes.

4.4 Cluster Configuration

All simulations are executed on The Aerospace Corporation's Fellowship [105] high performance computing cluster. Currently, Fellowship contains processors and servers totaling 1384 cores connected by Gigabit Ethernet to a dedicated Cisco Catalyst 6509 switch. The front-end servers provide job scheduling, shell access, directory service for user accounts, shared temporary file storage, network boot services, and backups. The cores run the Free Berkeley Software Distribution 6.2[6] operating system on a private non-routed network. Each simulation uses 250 cores from the current configuration shown in Table 4-5 via the Sun Grid Engine (SGE) batch queuing system.[7]

[6] http://www.freebsd.org, cited on July 24, 2008
[7] http://gridengine.sunsource.net/documentation.html, cited on July 24, 2008

Table 4-5: Fellowship nodes

Node Count	Type	Speed (GHz)	RAM (GB)	Core Count
64	Intel Xeon	2.0	1.0	512
32	Opteron 244	1.8	2.0	64
32	Opteron 244	1.8	2.0	64
24	Opteron 246	2.0	2.0	48
12	Opteron 246	2.0	2.0	24
32	Opteron 246	2.0	2.0	64
8	Opteron 252	2.6	2.0	16
116	Opteron 270	2.0	2.0	464
32	Opteron 275	2.2	2.0	128

The SGE uses a first-in-first-out scheduling policy to manage the shared Fellowship resource. When a job is submitted to the queue, if the quantity of cores required is available, execution begins immediately, if not, execution is delayed until the core quantity requirement can be met. When there are more cores available than required, SGE uses system load information on the machines to select the least loaded cores for processing.

In order to gain insight into the scheduler's core allocation strategy, a test was conducted wherein 4-, 8-, 16-, 32-, and 64-core requests were made. Each request was made 40 times to obtain a statistically meaningful sample of the overall frequencies with which SGE selects processor groups for executing job submissions. The distributions shown in Fig. 4-3b-f nearly match the Fellowship cluster core distribution illustrated in Fig. 4-3a.

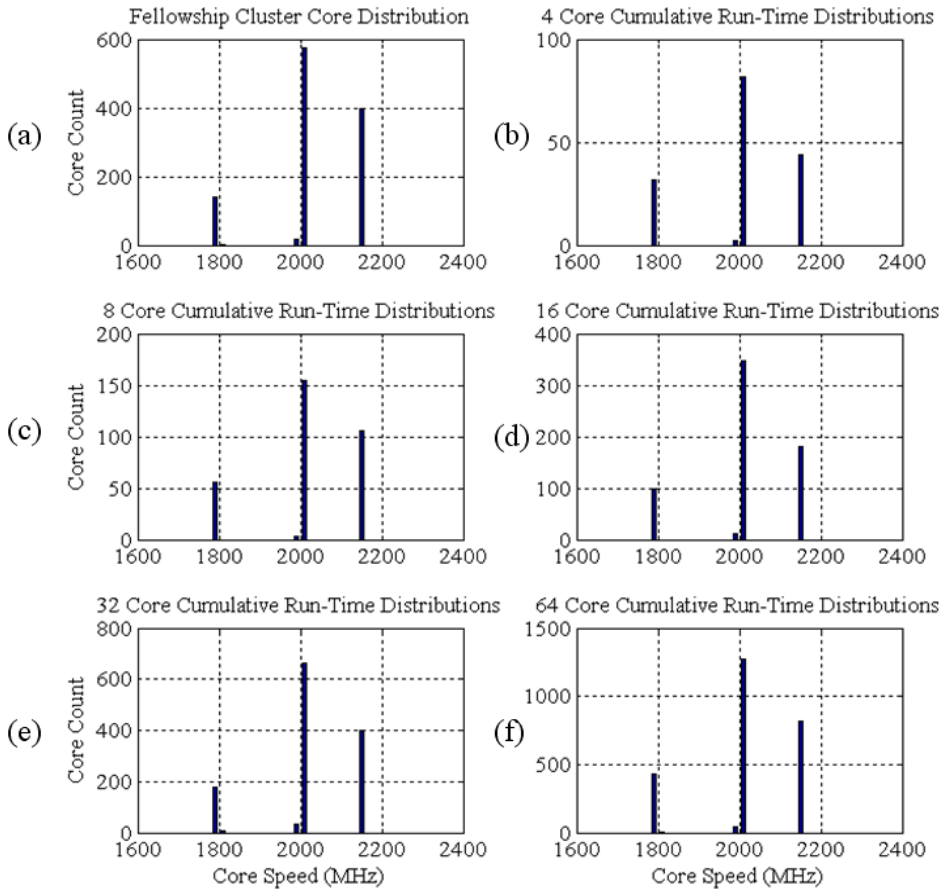


Figure 4-3: Fellowship core distribution test summary

This result indicates that most simulations will use a similar heterogeneous combination of cores for processing for each 250 core request.

Chapter 5

Pareto Hypervolumes and Decision Support

This Chapter presents the results from the individual case studies described in Chapter 4. Results include a discussion of the search dynamics and presentation of the Pareto-hypervolumes. The decision making process in a six-dimensional tradeoff landscape is presented and individual designs are selected for detailed examination and discussion.

5.1 Coverage Performance Baseline

The baseline coverage performance ($f_1(\bar{x})$ and $f_2(\bar{x})$ from Eqs. 4.3 and 4.4 respectively) of each degraded constellation before reconfiguration is shown in Table 5-1.

Table 5-1: Coverage performance prior to any reconfiguration, all cases

Cases-	ρ	$f_1(\bar{x})$ (min)	$f_2(\bar{x})$ (min)	Index(s) of Satellite, *s*, Lost
1a/c	6	1440.00	1440.00	24
1b/d	2	1439.99	1432.84	24
2a/c	6	1439.87	1402.53	21, 22, 23, 24
2b/d	2	1430.24	1345.47	21, 22, 23, 24

Cases-1a/c reveals that, even with the loss of the selected satellite, the GPS constellation is still able to provide four-fold coverage to every receiver with a PDOP better than 6.0

(above 5° minimum elevation). Since the hypervolumes sought maximize $f_1(\vec{x})$ and $f_2(\vec{x})$ and minimize $f_3(\vec{x})$ through $f_6(\vec{x})$, there can only be one Pareto-optimal solution for Cases-1a/c. The solution is the original 23-satellite constellation where no satellites are maneuvered resulting in the maximum possible values for $f_1(\vec{x})$ and $f_2(\vec{x})$ (1440.00 minutes) and minimum possible values for $f_3(\vec{x})$ through $f_6(\vec{x})$ (0.0). The result for Cases-1a/c presents an opportunity to test the pMOEA's ability to find a known result shown in Table 5-2.

Table 5-2: Pareto-hypervolume objective vector for Cases-1a/c

$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (m/s)	$f_5(\vec{x})$ (m/s)	$f_6(\vec{x})$
1440.00	1440.00	0.00	0.00	0.00	0
maximum values		minimum values			

For the remaining cases, no single solution defines each hypervolume because four-fold coverage to every receiver is not attainable without maneuvering. The coverage objective functions each provide a single value that quantifies the average or worst case performance over the receiver grid. It is instructive to understand how those single values map to their associated global coverage contours for each of the remaining three cases before any maneuvering takes place. The contours, shown in Fig. 5-1 to 5-3 for Cases-1b/d, Cases-2a/c, and Cases-2b/d, respectively, provide a baseline to compare the coverage performance of any reconfigured designs. Inside each color swatch, appearing

in the legend below the Robinson Earth projection, is the percentage of the total contour region with a FDVT between the printed values.

These contour plots show that the PDOP constrained accesses are *not* symmetrical between Northern and Southern hemispheres. This access behavior is due to the combination of the PDOP constraint and the loss of satellites from the nominal constellation. To better understand why, consider Fig. 5-4 where the global contour plot of the full 24-satellite GPS constellation with PDOP<1 and ξ>5º at epoch is presented. The contours for the nominal constellation present a mirror image between Northern and Southern hemispheres, so for this case, a hemispherical receiver grid provides the same global coverage information as a global grid. Now consider Fig. 5-5, where a full plane of the nominal constellation is removed. The resulting contours are no longer a mirror image of each other indicating that the Northern hemisphere cannot be used to *simulate* global coverage for the case studies considered in this research (a prior assumption from the experimental design Section 4.1.1). As a result, it is important to note that the non-dominated reconfigurations that follow are optimized to a receiver grid in the Northern hemisphere only.

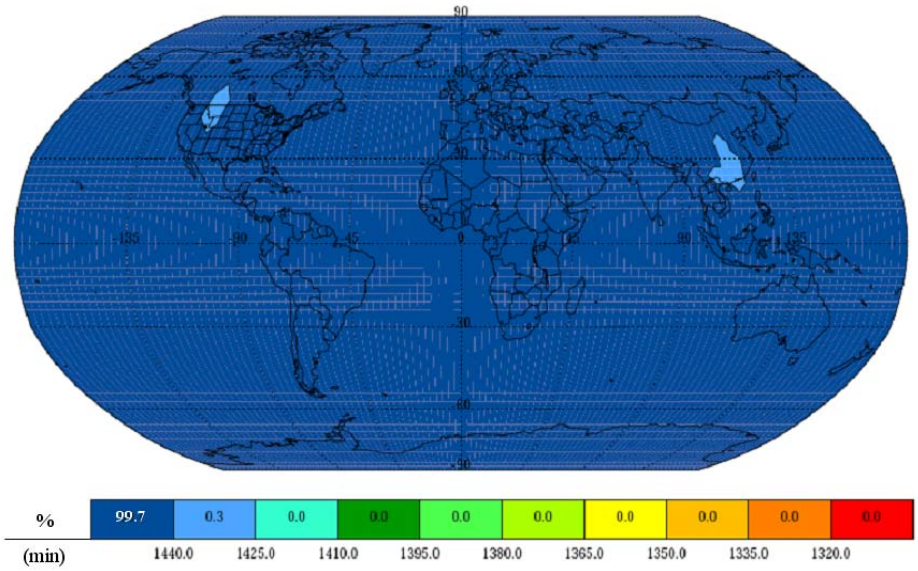


Figure 5-1: Cases-1b/d (one satellite lost, PDOP<2) DVT global contours before any reconfiguration

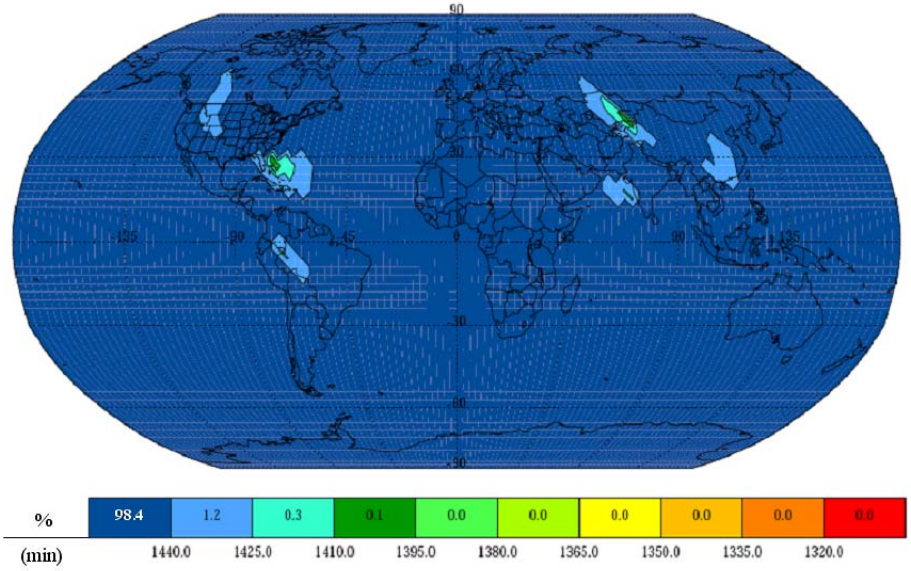


Figure 5-2: Cases-2a/c (one plane lost, PDOP<6) DVT global contours before any reconfiguration

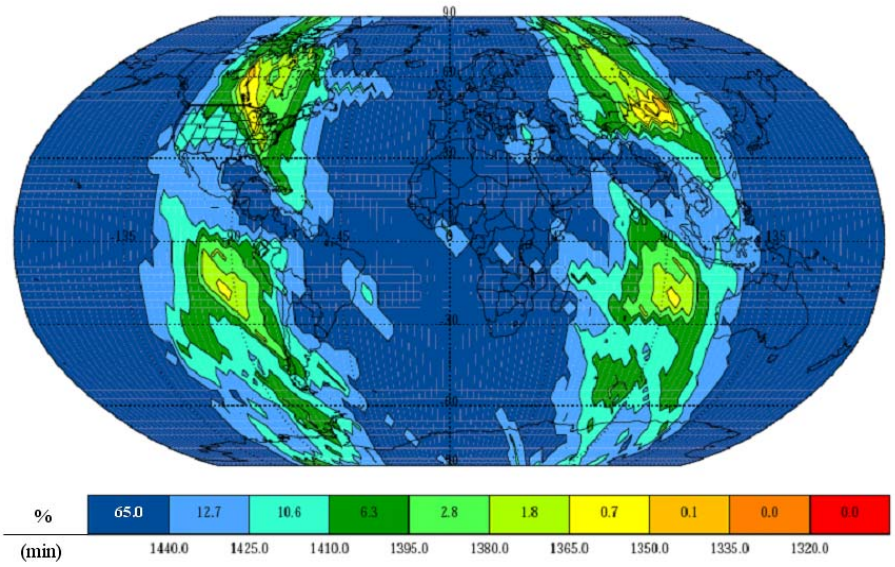


Figure 5-3: Cases-2b/d (one plane lost, PDOP<2) DVT global contours before any reconfiguration

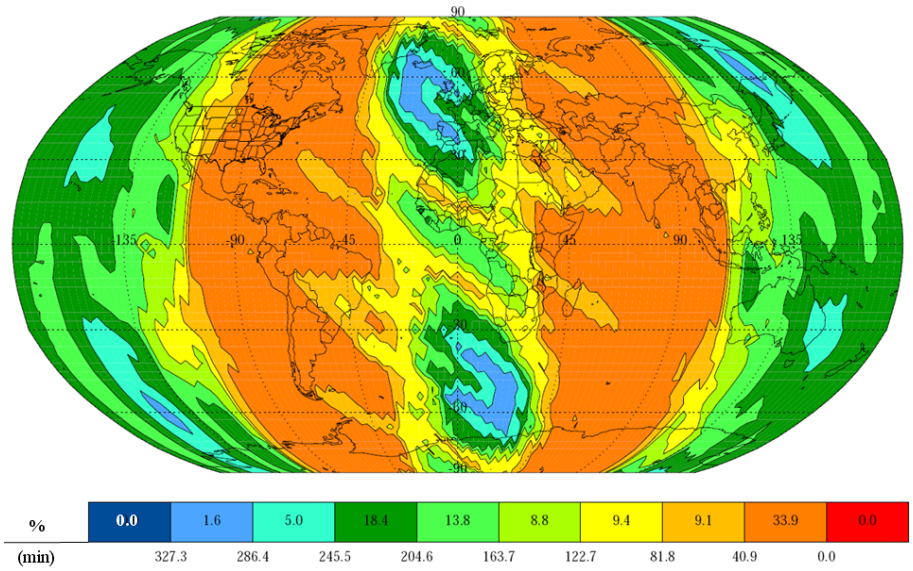


Figure 5-4: Global four-fold contour of original 24-satellite GPS constellation with PDOP<1 and ξ>5º at epoch

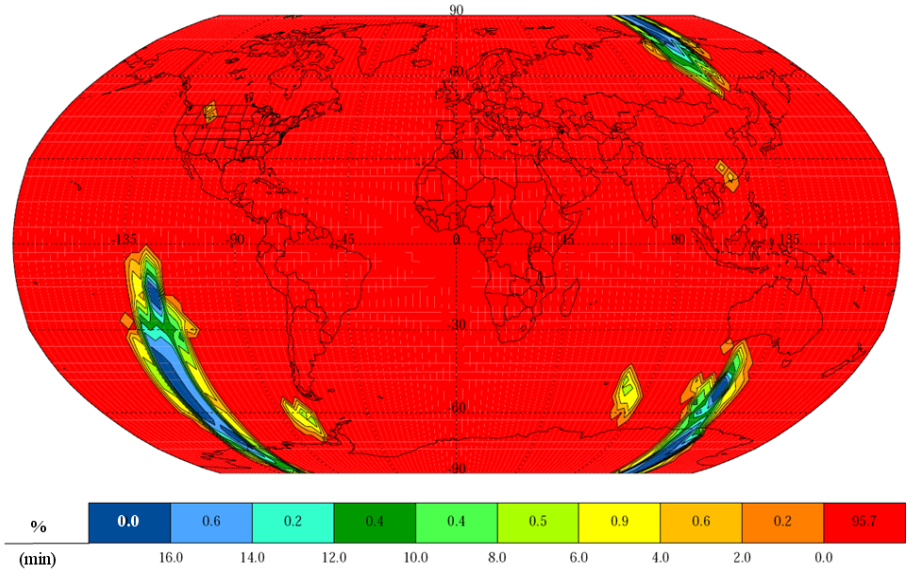


Figure 5-5: Global four-fold contour of the GPS constellation missing one plane with PDOP<1 and ξ>5º at epoch

5.2 Search Dynamics

The Pareto-hypervolumes for the k-dimensional (where k = 6) objective space contain the non-dominated solutions present in the archive at optimization run termination. The Fellowship cluster enforces maximum run durations of 95 hours for any set of processing resources allocated for a particular job (this limit is imposed because Fellowship is a shared parallel computing resource). In many cases, for a variety of reasons (node failures, etc.), the job is terminated before the 95-hour period has elapsed. The check-pointing mechanism allows the job to pick up where it left off. The decision to terminate each run is a pragmatic one. Periodically, at each check-point or at the time of a node failure, if a sufficient solution quality has been reached and the computational

demands associated with further search are not warranted, then the optimization is halted. This philosophy driving the termination criteria is known as satisficing [107] (typically better than has been previously demonstrated), and is necessary given the computational demands of this research.

One way to gain insight into the impact that the assumptions and constraints of each reconfiguration case have on the characteristics of the fitness landscape is to observe the archive size dynamics leading up to run termination. These are shown for Cases-1a/c-ideal and Cases-1a/c-op (one satellite lost with PDOP<6) in Fig. 5-6 where each data point is printed at ten generation intervals.

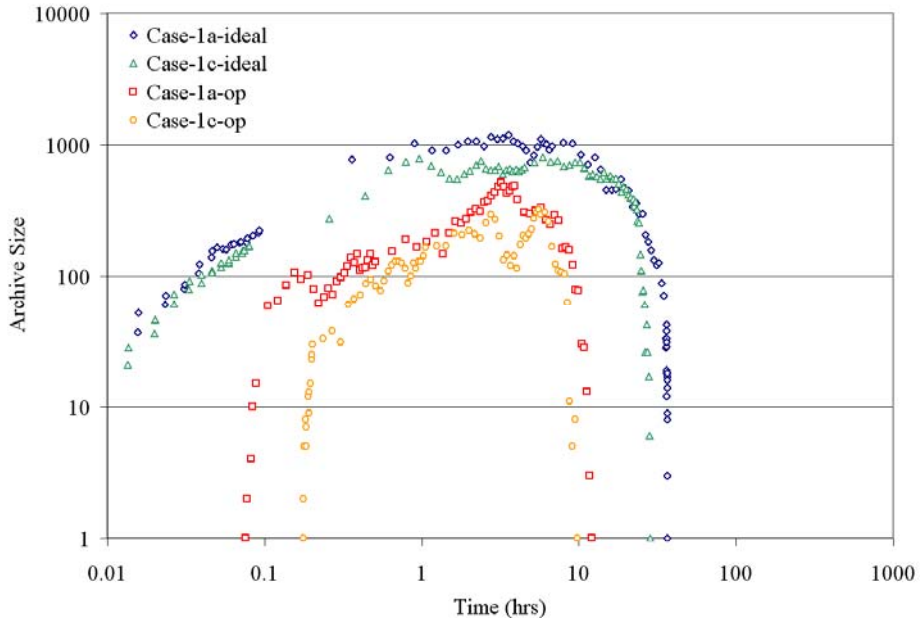


Figure 5-6: Archive dynamics, one satellite lost, PDOP<6, all cases

In every case, the LC-ε-NSGA-2 was able to successfully converge to the single known Pareto-optimal solution. The operational cases initially have no archive members because only feasible solutions (e.g., those who do not violate the propellant constraints) are allowed to enter the archive. The traces for Case-1a-op and Case-1c-op also grow and shrink more frequently than those with ideal assumptions indicating that considerable multimodality was introduced into the fitness landscape through the inclusion of operational constraints. For the idealized cases of Fig. 5-6 (and Figs. 5-8, 5-9, 5-10), shortly after 0.1 hours, the first interconnected run finishes and population reinvigoration occurs (i.e., time continuation as described in Section 3.1.3). The clear gap between the blue and green markers after 0.1 hours is due to the extended function evaluation time required to evaluate each generation of the much larger archive at the start of the next interconnected run.

The time to convergence for Case-1a-ideal/op, where K is common among all satellites, and Case-1c-ideal/op, where K is independent for each is shown in Fig. 5-7 (recall D is the number of decision variables).

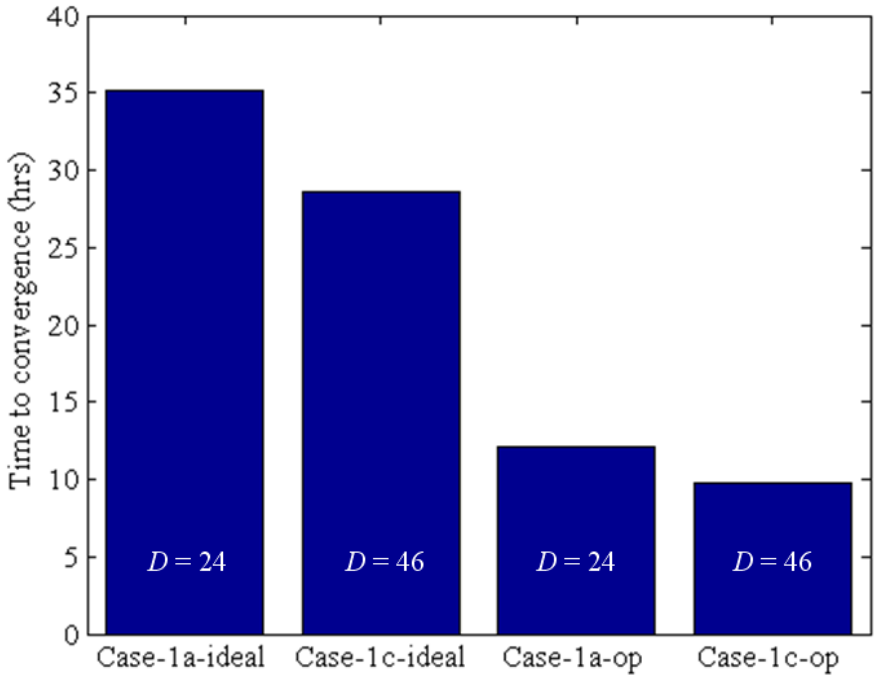


Figure 5-7: Convergence time for Cases-1a/c-ideal and Cases-1a/c-op

The introduction of operational constraints cut the convergence time to one third of those cases that had access to unlimited energy. A glance at Fig. 5-6 reveals why. By imposing propellant limits, the number of potential solutions that can make up the archive as search progresses are considerably less than those with idealized assumptions.

When the number of decision variables increases from 24 to 46, convergence occurs more rapidly in both the ideal and operational cases. Allowing K to be independent created many more potential paths in the search space, so the result might seem counterintuitive since the chromosome increased in complexity, but required less search time. This observation, however, is better understood by considering the impact

that a common vs. independent K has on the fitness landscape. Requiring K to be common among satellites maneuvered created *feasibility islands* wherein a large portion of the search space could not be accessed. Releasing K, such that it is independent for each satellite maneuvered *bridged* the feasibility gaps allowing for more rapid convergence, despite the increase in chromosome size.

The search dynamics for all remaining cases, where the Pareto-hypervolumes are not known *a priori*, are illustrated in Figs. 5-8 through 5-10. Since the final hypervolumes are not known for any of the remaining cases, any sweeping claims regarding convergence will be avoided; however, all solutions at run termination were checked to ensure that a satisfactory quality was attained. The results presented in this section will motivate the suggestion of a new termination criterion in Section 5.4 that will add rigor to the *satisficing* philosophy.

By counting the number of times that the trace for the archive size changes direction, it is possible to gain insight into the modality of the search space created by the assumptions. As with the previous cases, every case of Figs. 5-8 through 5-10 shows that the introduction of operational considerations (red and orange data markers) creates multi-modality in the search space that was not present in the cases with access to unlimited energy (blue and green data markers). In both Figs. 5-6 and 5-8, one satellite was lost, but in Fig. 5-8 the PDOP constraint was tightened. Recall the effect of reducing the PDOP to 2 is to create more access gaps in the coverage, presumably making the problem harder to solve. The search traces of Fig. 5-8 extend well beyond those of Fig. 5-6 supporting this claim. As before, Fig. 5-8 also shows that the introduction of

operation assumptions allow for earlier run termination than the cases whose satellites have access to unlimited energy.

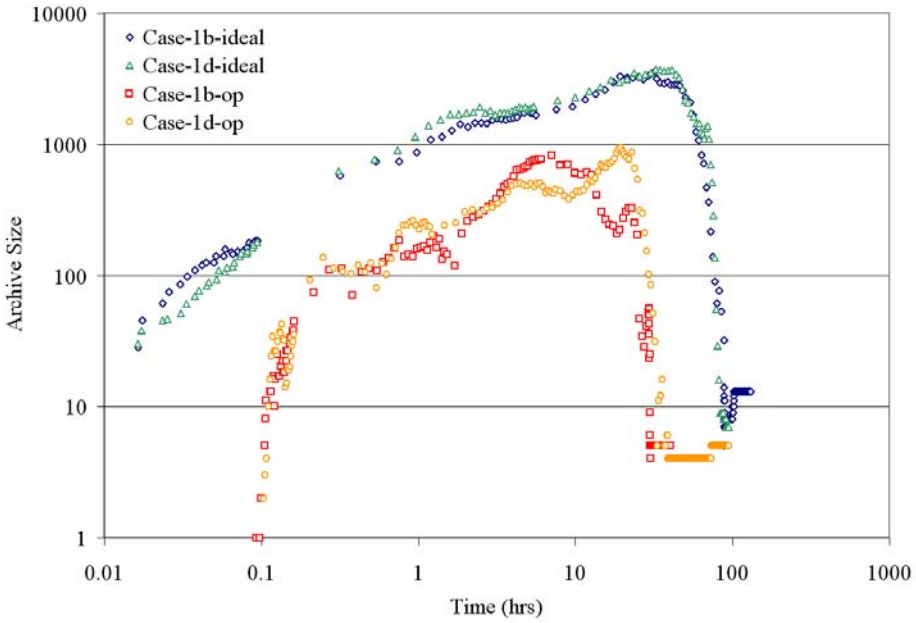


Figure 5-8: Archive dynamics, one satellite lost, PDOP<2, all cases

In Figs. 5-9 and 5-10, the search traces for the loss of an entire plane (Case-2) are shown. The size of the archives and time at run termination in Fig. 5-10 are an order of magnitude greater than those of Fig. 5-9 that are also an order of magnitude greater than Fig. 5-8. This observation indicates that the loss of an entire plane and the tightening of the PDOP constraint make the problem exponentially more difficult to solve. The reason for this is likely due to the impact that reducing the size of the constellation and tightening of the PDOP constraint has on the access patterns. A brief glance at the contour plots of Figs. 5-1 through 5-3 shows that the coverage becomes progressively

worse, in an unsymmetrical manner, both as more satellites are subtracted from the constellation and as greater demands are placed on PDOP performance. As the access gaps increase, so do the number of potential reconfiguration alternatives. Also note that, while the number of decision variables decreases when an entire plane is subtracted from the constellation, the problem became more difficult to solve. Further, data markers representing the operational cases (red and orange) of Figs. 5-9 and 5-10 begin to intermingle with the idealized cases (green and blue data markers) indicating that it is no longer possible to claim that operational assumptions make the problem easier to solve. In fact, the opposite is true because the operational cases of Fig. 5-10 appear to be making progress even after 500 hours of search.

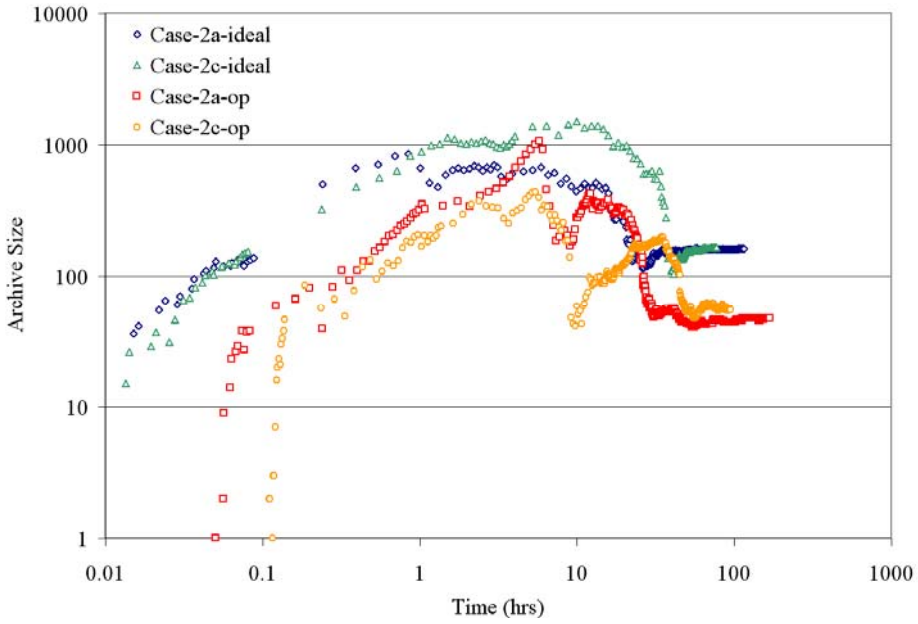


Figure 5-9: Archive dynamics, one plane lost, PDOP<6, all cases

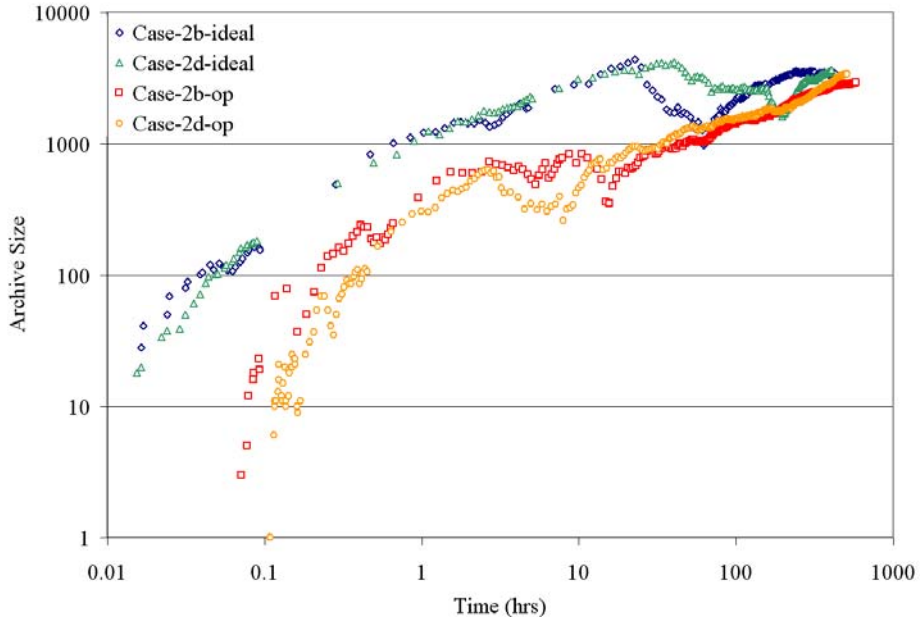


Figure 5-10: Archive dynamics, one plane lost, PDOP<2, all cases

Any problem that requires search for nearly a month (even with 250 cores) will not necessarily enable a rapid turnaround on the reconfiguration decision. Even though search progress was made, Case-2b-op and Case-2d-op fit into this category and represent exceptions to the run termination criterion outlined earlier. Search on these cases was halted near 500 hours because they had used an exorbitant amount of processing resources in the shared parallel computing environment. It was previously mentioned that increasing the size of epsilon values has the potential to increase convergence speed. Later in this chapter, Case-2d-op will be revisited with an epsilon vector that is coarser than the original used throughout the majority of this research. At that point, it will be possible to compare the hypervolumes to determine what, if anything, was lost or gained in solution quality for the potential reduction in search time.

5.3 Case Studies

Of the remaining reconfiguration case studies outlined in Section 4.2, half of them (Case-1b, Cases-2a/b) required K to be common among all satellites maneuvered. In the previous section, these cases along with the others (Case-1d, Cases-2c/d) presented an opportunity to make observations regarding the impact of chromosome size on search progress. Since, in reality, there would generally be no reason to limit K to be common among all satellites maneuvered, any subsequent presentation of the resulting hypervolumes and their designs for Case-1b and Cases-2a/b are deferred to Appendix G. Those results are included for completeness, but the decision-making process for the figures presented in Appendix G is identical to what follows for Case-1d and Cases-2c/d (reference Table 5-3 for a review of these cases). The following sections will consider a total of six scenarios and are organized into four sections (5.3.1 to 5.3.4) according to the number of satellites lost and the idealized or operational assumptions applied.

Table 5-3: Review of decision vectors for independent K_s cases

Case-	ρ	Decision Vector	Decision Variables, D	Index(s) of Satellite, s, Lost
1d-ideal	2	$\vec{x} = [M_1 \quad M_2 \quad ... \quad M_{23} \quad K_1 \quad K_2 \quad ... \quad K_{23}]^T$	46	24
2c-ideal	6	$\vec{x} = [M_1 \quad M_2 \quad ... \quad M_{20} \quad K_1 \quad K_2 \quad ... \quad K_{20}]^T$	40	21,22,23,24
2d-ideal	2	$\vec{x} = [M_1 \quad M_2 \quad ... \quad M_{20} \quad K_1 \quad K_2 \quad ... \quad K_{20}]^T$	40	21,22,23,24
1d-op	2	$\vec{x} = [M_1 \quad M_2 \quad ... \quad M_{23} \quad K_1 \quad K_2 \quad ... \quad K_{23}]^T$	46	24
2c-op	6	$\vec{x} = [M_1 \quad M_2 \quad ... \quad M_{20} \quad K_1 \quad K_2 \quad ... \quad K_{20}]^T$	40	21,22,23,24
2d-op	2	$\vec{x} = [M_1 \quad M_2 \quad ... \quad M_{20} \quad K_1 \quad K_2 \quad ... \quad K_{20}]^T$	40	21,22,23,24

5.3.1 One Satellite Lost, Idealized Assumptions

To briefly review, with Case-1c-ideal, the single Pareto-optimal solution was found after 1,095,071 function evaluations and 28.6 hours of search on 250 processing cores. Case-1d-ideal, where the PDOP constraint was tightened to 2, achieved the seven solutions, whose objective vectors are shown in Table 5-4, after completing 3,357,055 function evaluations during the first 95-hour checkpoint period.

Table 5-4: Hypervolume objective vectors, Case-1d-ideal (one satellite lost, PDOP<2)

Design	$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (m/s)	$f_5(\vec{x})$ (m/s)	$f_6(\vec{x})$
1	1439.99	1437.00	1.00	6.58	0.00	1
2	1439.99	1435.00	1.00	3.42	0.00	1
3	1439.98	1432.84	0.00	0.00	0.00	0
4	1439.99	1437.00	1.50	4.39	0.00	1
5	1440.00	1440.00	2.00	4.46	0.00	1
6	1440.00	1440.00	1.00	8.91	0.00	1
7	1440.00	1440.00	0.99	10.52	0.00	1

While presenting the objective vectors in table format is appropriate for small data sets, it is cumbersome for many more than 20 designs. One way to visualize hundreds or even thousands of designs simultaneously is to use a glyph plot where data markers are assigned different attributes for each dimension of the problem. All hypervolume glyph plots in the pages that follow associate each objective function with the attributes shown in Table 5-5.

Table 5-5: Plotting attributes for all glyph hypervolumes

Objective	Abbreviation	Plotting Attribute
$f_1(\vec{x})$	A-DVT	marker size
$f_2(\vec{x})$	WCP-DVT	z-axis
$f_3(\vec{x})$	TTOF	color
$f_4(\vec{x})$	Worst ΔV or Worst m_u	y-axis
$f_5(\vec{x})$	ΔV Var or m_r Var	cone direction
$f_6(\vec{x})$	Man	x-axis

Three of the objectives ($f_2(\vec{x}), f_4(\vec{x})$, and $f_6(\vec{x})$) are plotted in $\Re^3$, while the remaining three are illustrated with the size ($f_1(\vec{x})$), direction ($f_5(\vec{x})$), and color ($f_3(\vec{x})$) of the conical marker. A small blue cone pointing downward indicates a design that is closer to the lower end of the plotting range. Since the goals are to minimize TTOF and the sum of the ΔV variance (or m_r variance) and maximize the A-DVT, the decision maker seeks designs that are represented by downward pointing, large markers colored towards the blue end of the color scale. For the remaining three objectives, the locations of the desirable designs are a function of the orientation of the $\Re^3$ volume. For the hypervolume of Case-1d-ideal shown in Fig. 5-11, the designs in the lower left corner towards the back of the $\Re^3$ volume are best, but the reader should not assume that the orientation will remain consistent from figure to figure. In some cases, rotation of the volume is necessary for analysis and, as such, the reader should take a moment with each glyph plot to orient themselves to both the scale for each objective function and positioning of the $\Re^3$ volume.

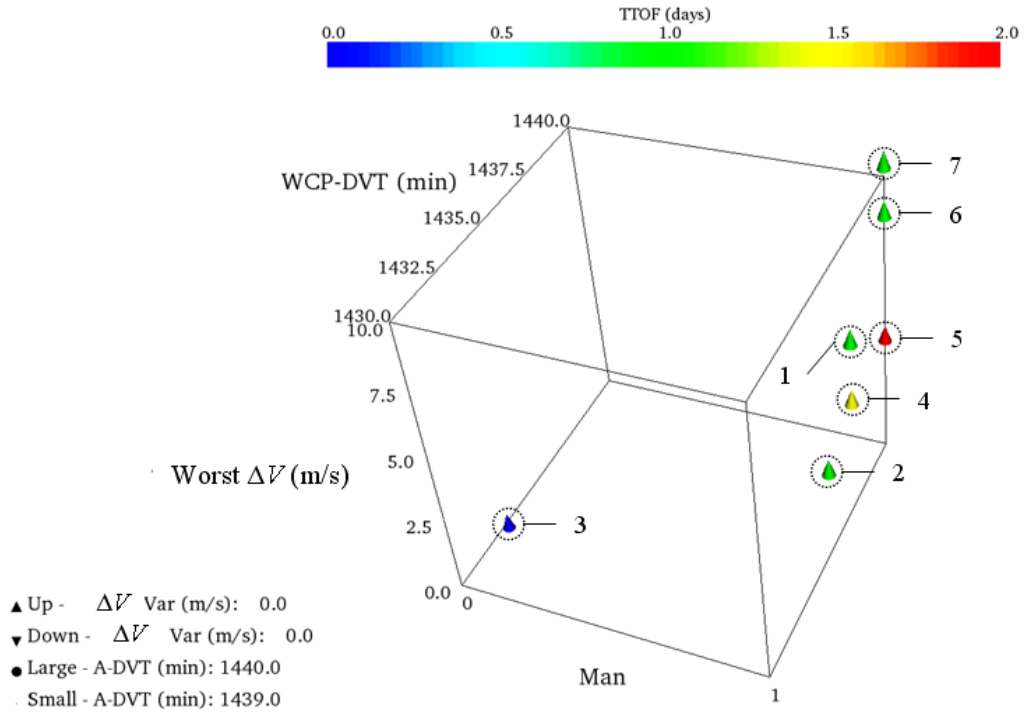


Figure 5-11: Glyph hypervolume at run termination, Case-1d-ideal (one satellite lost, PDOP<2)

As with Case-1c-ideal, the solution where no satellites are maneuvered appears in the hypervolume for Case-1d-ideal as Design 3 from Table 5-4. In contrast to this design for 1c, Design 3 does not achieve perfect coverage because of the tightened PDOP constraint. Of the remaining designs in Table 5-4, all maneuver only one satellite and three designs (5, 6, and 7) attain continuous coverage to all receivers. If it is critical that the constellation restore continuous four-fold coverage, the obvious selection is Design 5 because it costs the least amount of energy and takes only one extra day to accomplish the reconfiguration relative to the higher energy solutions. Since the remaining solutions

do not offer any dramatic improvement in coverage performance, and if four-fold coverage is not essential, the hypervolume could be used to justify the decision to simply accept the loss and operate with the current configuration.

While the hypervolumes provide a six-dimensional overview of the objectives in Λ, the individual decision variables must be examined to determine which satellites are rephased and by how much. In Fig. 5-12, each design is given a vertical slice of the plot and, if maneuvered, the change in phase (the integer number of phasing orbits, K_s, plus its phase angle, $\alpha_s/360°$) appears as a colored block. For small archives, like that of Fig. 5-12, the design number is shown inside of the color block for reference purposes only. The hypervolume designs are sorted by number of satellites maneuvered in ascending order along the horizontal axis delineated by black vertical lines. The white lines that horizontally cross the plots are intended to provide delineation between the individual planes of the constellation. During any preceding discussion of these decision variable figures, satellites 1 to 4 are referred to as occupying plane 1, satellites 5 to 8 as occupying plane 2, and so on.

A counterintuitive observation from Fig. 5-12 is that not a single satellite is reconfigured from plane 6 where the satellite was lost. Instead, the majority of designs rephase satellites 9 and 10 in the plane *opposite* to plane 6. This indicates that, even with idealized conditions, the more intuitive approach of redistributing the satellites symmetrically within the plane where the loss occurred does not produce a non-dominated choice. The lesson here is to avoid the tendency towards symmetry in problems involving constellation design, especially when the problem's assumptions deviate from Walker's [11] purely geometric, idealized case.

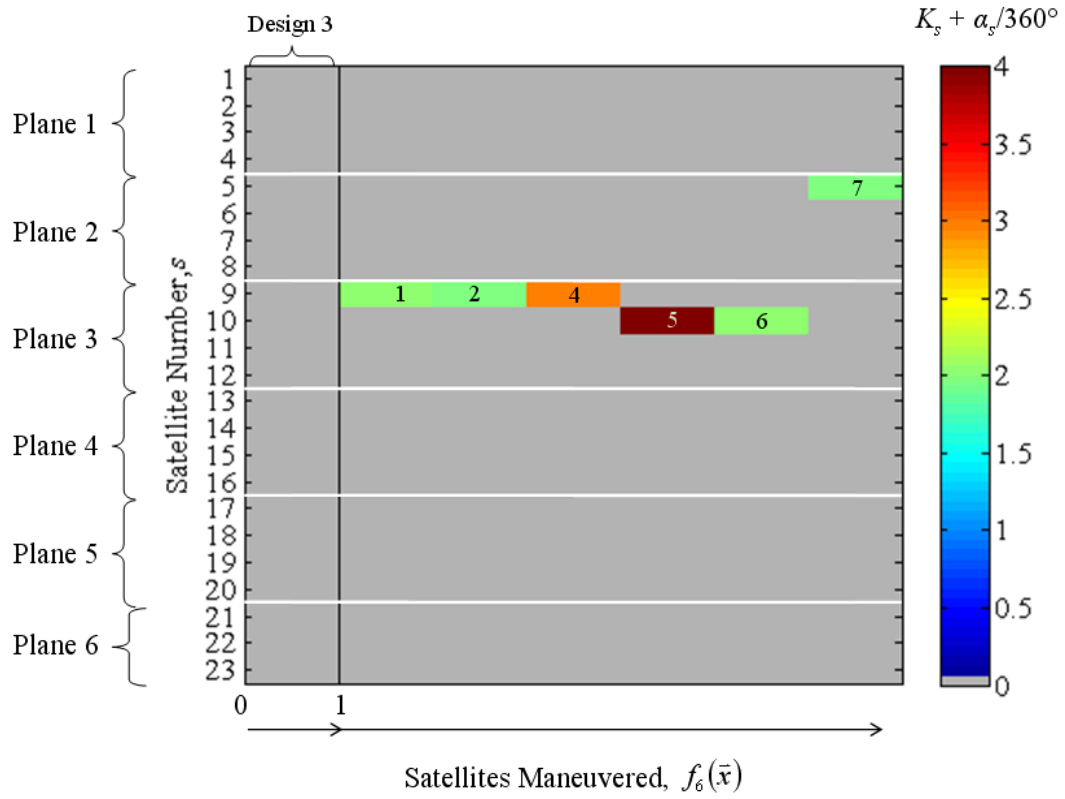


Figure 5-12: Hypervolume decision variable plot, Case-1d-ideal (one satellite lost, from plane 6, PDOP<2) independent phasing orbits, K_s, plus change in phase, α_s, for each epsilon-non-dominated design, sorted by $f_6(\bar{x})$ ascending

5.3.2 One Satellite Lost, Operational Assumptions

The operational assumptions, originally detailed in Table 4-1, incorporate engine and satellite mass characteristics and introduce propellant constraints. Figure 5-13 shows the initial state of each satellite's propellant tank at the time that the constellation

experiences the loss. The propellant-gage-like representations show the propellant already depleted before any maneuver is performed (blue portion, m_d) and the propellant available for maneuvering (red portion, m_r). Recall from Chapter 4 that m_r is drawn randomly from a uniform distribution with amplitude equal to 25 kg so the percentages displayed are the fraction of that total. Figures of this type are used to illustrate the actual propellant expenditure for each satellite for any given non-dominated design.

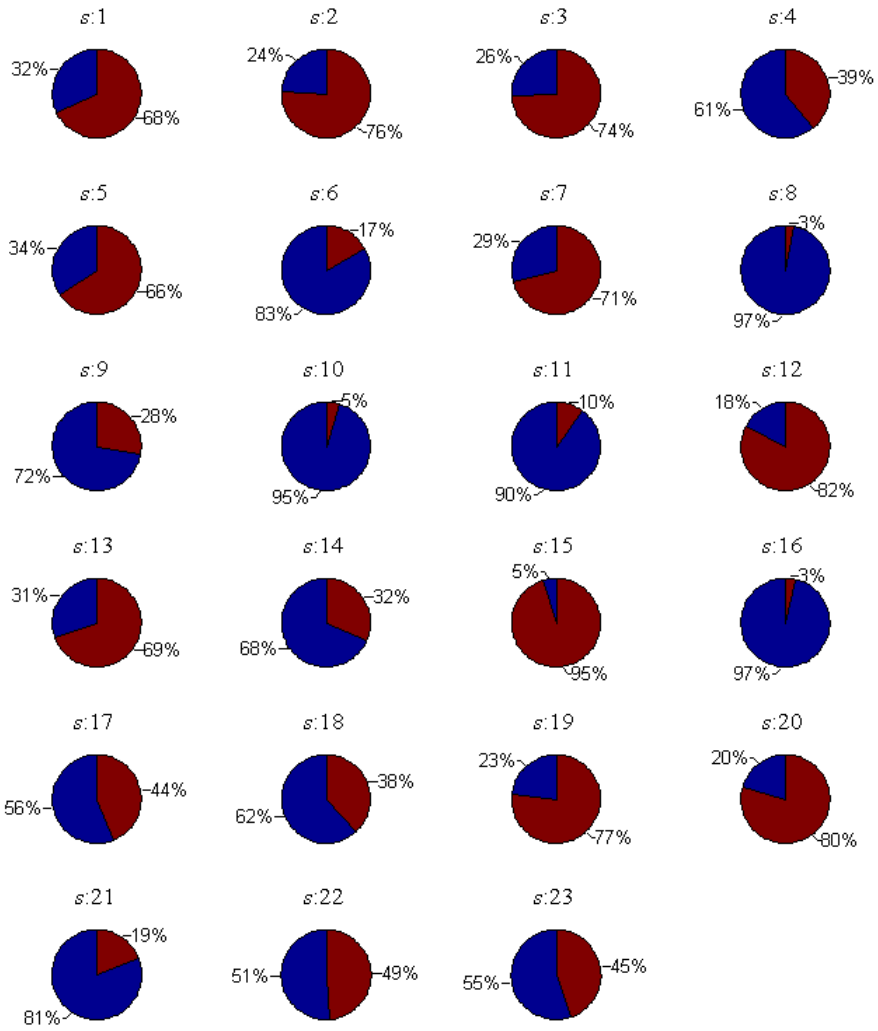


Figure 5-13: Initial propellant state of the constellation
blue: % mass depleted, red: % mass remaining

The single Pareto-optimal solution for Case-1c-op was found in 9.8 hours on 250 processing cores after 344,311 function evaluations. Case-1d-op, with the tightened PDOP constraint, achieves five solutions that have the objective vectors shown in Table 5-6 after a full checkpoint period of 95.0 hours has elapsed during which 3,274,759 function evaluations were completed. Even though both the ideal and operational Case-

1d completed approximately the same number of function evaluations in the same wall clock period, Case-1d-ideal only completed several hundred (284) interconnected runs, much fewer than the thousands (5424) of the operational case. This suggests that the hypervolume shown in Fig. 5-14 was found much earlier in the 95.0-hour search period indicating that the introduction of operational assumptions made Case-1d easier to solve.

Table 5-6: Hypervolume objective vectors, Case-1d-op (one satellite lost, PDOP<2)

Design	$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (kg)	$f_5(\vec{x})$ (kg)	$f_6(\vec{x})$
1	1440.00	1440.00	0.99	5.06	0.00	1
2	1440.00	1440.00	1.49	3.37	0.00	1
3	1439.9877	1432.84	0.00	0.00	0.00	0
4	1439.9958	1437.00	0.99	3.38	0.00	1
5	1439.9988	1438.46	1.00	4.04	0.00	1

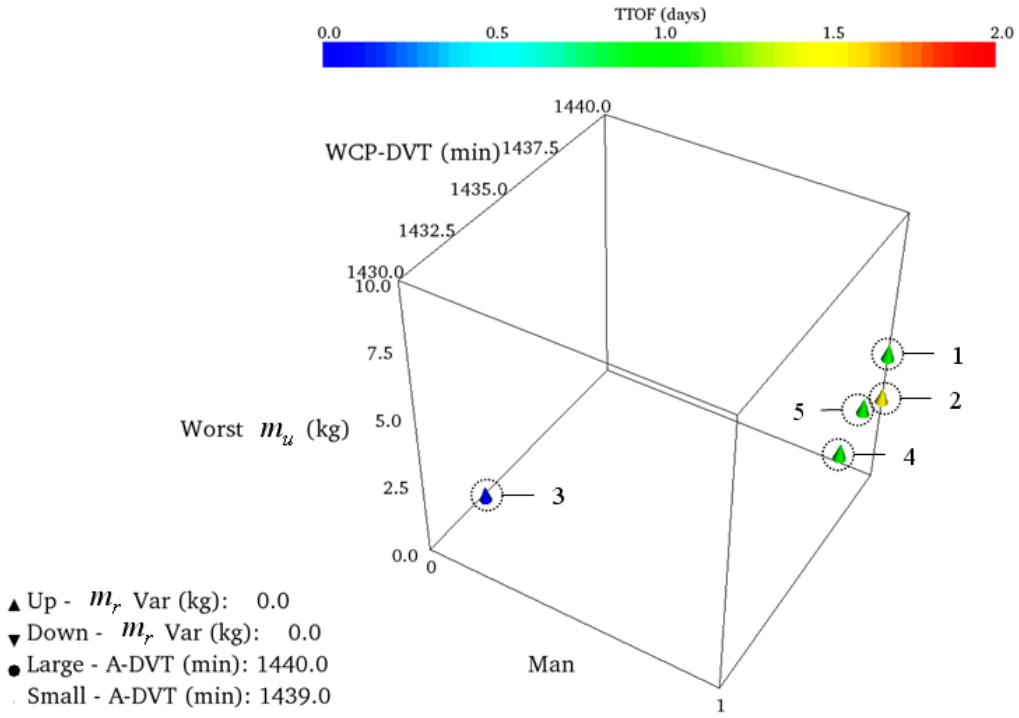
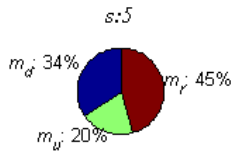
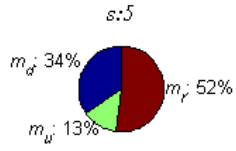
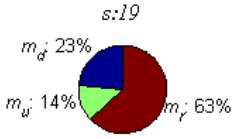
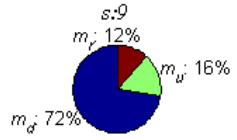


Figure 5-14: Glyph hypervolume at run termination, Case-1d-op (one satellite lost, PDOP<2)

Note that Design 3 (the solution where no satellites are maneuvered) still meets both the A-DVT and WCP-DVT coverage requirements per the specification; however, there are several designs that provide continuous coverage to the receiver grid, and one obvious choice (Design 2) from a minimum propellant perspective. However, since each satellite has a different amount of propellant, it is not enough to use only Table 5-6 or Fig. 5-14 to make a decision. Instead, the propellant-state information is introduced in Table 5-7 to further aid the analysis. The green colored sections indicate the percentage of propellant used as a result of the maneuver, m_u (note that because of rounding in the graphic, it is possible that the percentages do not appear to add up to 100%).

Both Designs 1 and 2, due to their continuous coverage, compete for the most preferred choice. Table 5-7 shows that Design 2 leaves 52% of the propellant remaining post-maneuver in contrast to the more rapid reconfiguration described by Design 1 leaving 45% with both options rephasing the same satellite. If continuous coverage is not critical, then Designs 4 and 5 can be considered, but it is likely that Design 5 would be removed from consideration since the reconfiguration only leaves 12% of the propellant remaining on satellite 9 for future maneuvering.

Table 5-7: Propellant-gage plot (reduced to show only those maneuvered) hypervolume Designs 1, 2, 4, and 5, Case-1d-op (one satellite lost, PDOP<2)

Design	Propellant Gage
1	
2	
4	
5	

By comparing the decision-variable plot of Fig. 5-15 to that of Fig. 5-12, it is possible to determine the impact that the introduction of operational assumptions has on the satellites that are selected for maneuvering. In Fig. 5-15, a different pattern emerges than was previously observed. Here, there is no particular trend between both the most common satellite maneuvered and the planes in which they are rephased. Further, the satellites that have the most available propellant (e.g., satellite 15) or the highest specific impulse and/or propellant combination (e.g., satellite 12) are not maneuvered. This observation leads to another important point. When deciding which satellites to maneuver, propellant remaining, engine impulse, and satellite mass should not be considered in isolation (e.g., maneuvering satellite 15 would produce a result that is dominated and hence inferior to any of the choices shown in Table 5-7).

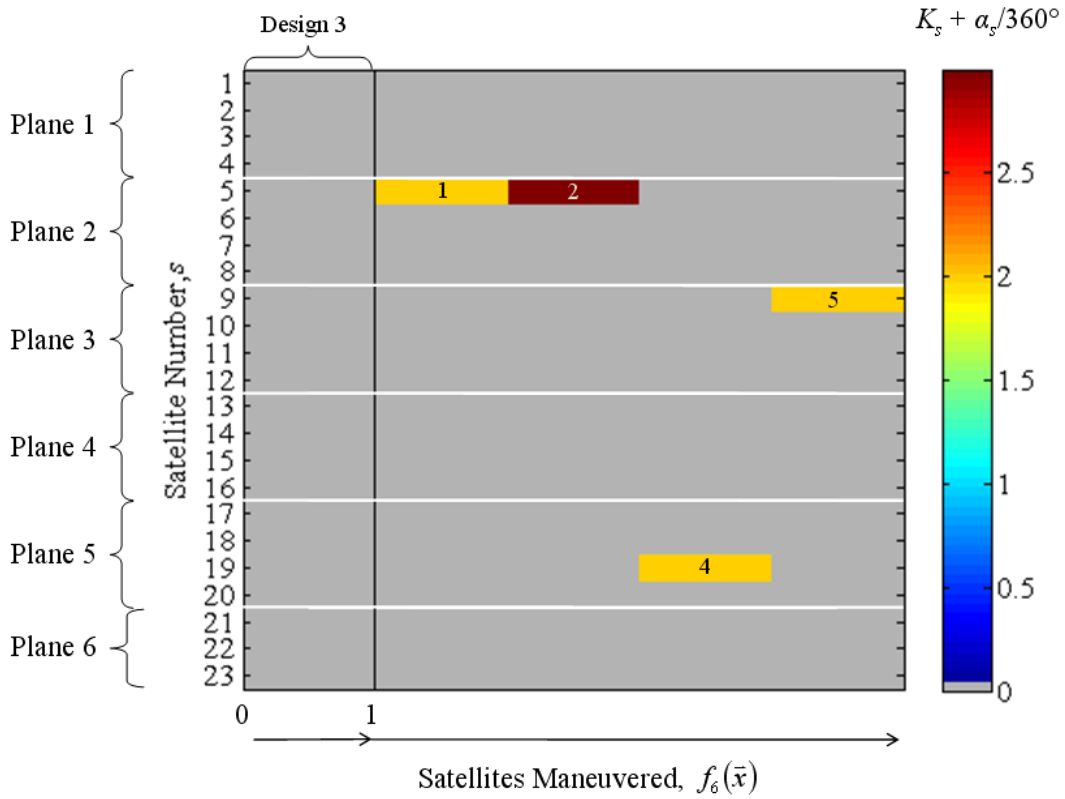


Figure 5-15: Hypervolume decision variable plot, Case-1d-op (one satellite lost, PDOP<2) independent phasing orbits, K_s, plus change in phase, α_s, for each epsilon-non-dominated design, sorted by $f_6(\vec{x})$ ascending

5.3.3 One Plane Lost, Idealized Assumptions

5.3.3.1 Access Constrained to PDOP ≤ 6

When an entire plane is lost, recall that, with a PDOP constraint of 6, the coverage performance improvements offered by the hypervolume designs must offer improvements on the 'do-nothing' solution, which has an A-DVT and WCT-DVT of

1439.87 and 1402.53 minutes, respectively. After 3,039,223 function evaluations (101 interconnected runs) and 95.0 hours of search on 250 processing cores, Case-2c-ideal found the 165 non-dominated solutions shown in Fig. 5-16. One of the most noticeable differences between this hypervolume and those of the previous cases is that non-dominated solutions exist for maneuvering zero satellites up to four satellites. Additionally, the number of solutions available to a DM is greatest when maneuvering one satellite, and choices decrease as the number of satellites maneuvered increases.

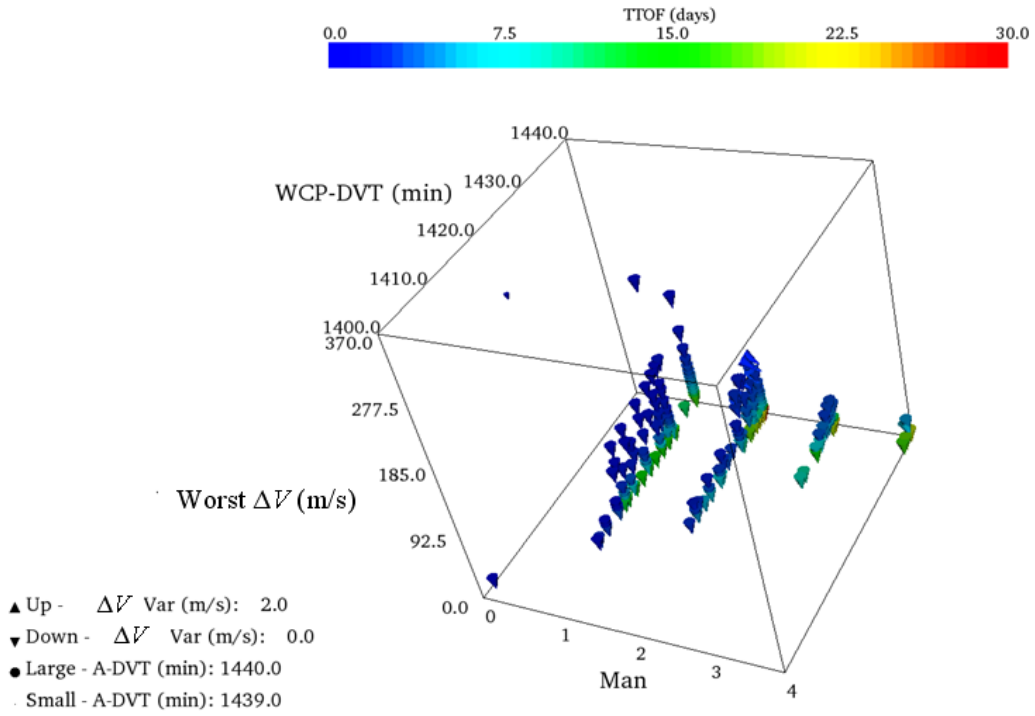


Figure 5-16: Glyph hypervolume at run termination, Case-2c-ideal (one plane lost, PDOP<6)

The decision-variable plot for the hypervolume shown in Fig. 5-16 is given in Fig. 5-17 with the designs sorted in ascending order according to the number of satellites that undergo maneuvering. No clear pattern emerges, however, when one satellite is maneuvered, most often only satellites 4, 6, and 16 are rephased. As more than one satellite is reconfigured, satellites 4, 9, 12, and 18 join those rephased. It can also be observed that there are many satellites that never undergo reconfiguration for any non-dominated design.

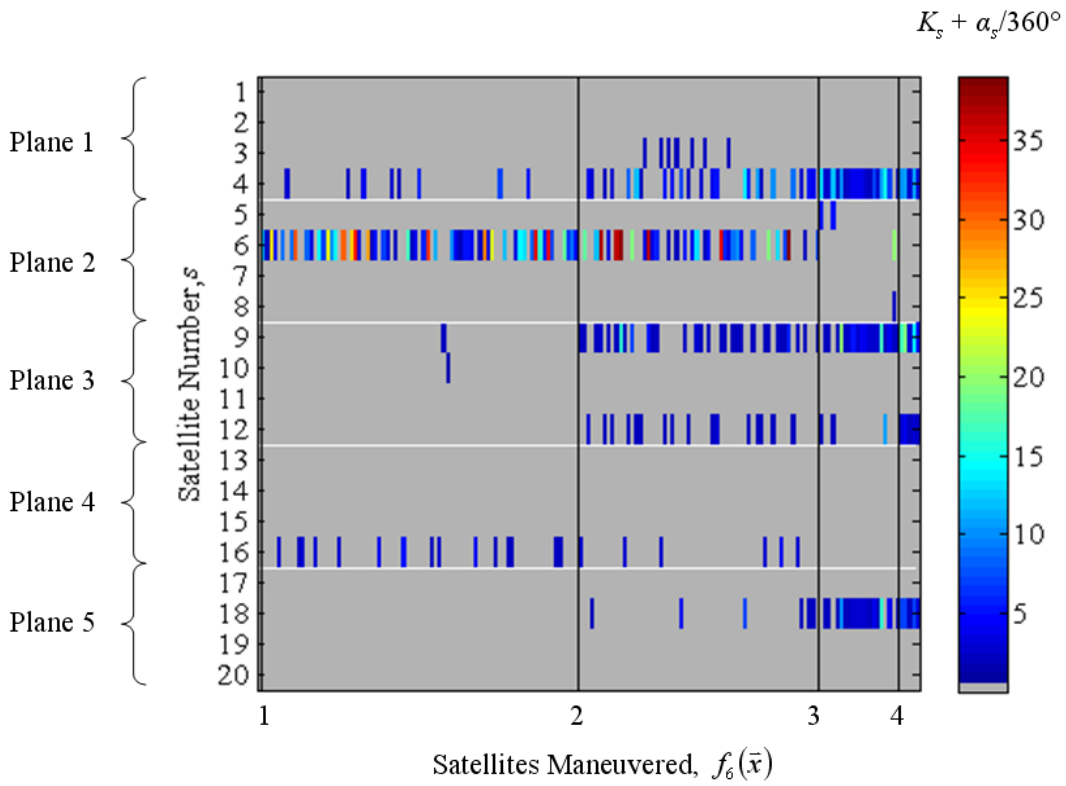


Figure 5-17: Hypervolume decision variable plot, Case-2c-ideal (one plane lost, plane 6, PDOP<6) independent phasing orbits, K_s, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\vec{x})$ ascending

A hypervolume with 165 solutions can be challenging to a DM. In the previous cases, the relatively small non-dominated sets could be easily explored and decisions made. This is not the case with the current hypervolume and those that follow. Ultimately, a single decision must be made and the large number of designs synthesized in a reasonably short period of time adding to the difficulty of the process. Towards this end, a *design-by-shopping* paradigm is used to overcome decision challenges.

Balling [108] classifies the design-by-shopping paradigm as an *a posteriori* articulation of preferences used to solve multi-objective optimization problems. In this research, we will modify the definition slightly to read: an *a posteriori* articulation of preferences on a set of non-dominated designs used to identify a set of *most preferred* solutions for the purposes of decision making. Stump *et al.* [109] describe an approach for calculating the most preferred design among a multi-dimensional set of data given a DM's preference structure. The approach constructs a weighting vector, $\vec{W}$, that has elements associated with each individual objective function. The dot product of $\vec{W}$ and the norm of each objective in $\vec{f}$, (using the maximum and minimum values of each objective to calculate the norm),

$$\vec{V} = \vec{W} \cdot \left\langle \left\| \vec{f}_1 \right\| \quad \left\| \vec{f}_2 \right\| \quad \cdots \quad \left\| \vec{f}_6 \right\| \right\rangle, \tag{5.1}$$

is constructed in the hyperspace defining a preference direction. The first non-dominated design that intersects the hyper-plane normal to the preference direction is the most

preferred. For example, consider the two-dimensional minimization problem shown in Fig. 5-18, where the objectives f_1 and f_2 are minimized. Assuming both objectives are equally important produces a preference direction at a 45° angle to the horizontal axis passing through the origin. The most preferred design is then identified by constructing a vector perpendicular to the preference direction at the utopia point ({0,0} in this case). That vector is then moved backwards, against the preference direction, until it encounters the first non-dominated design. This solution, identified by the green marker in Fig. 5-18, is the most preferred design for the associated preference structure.

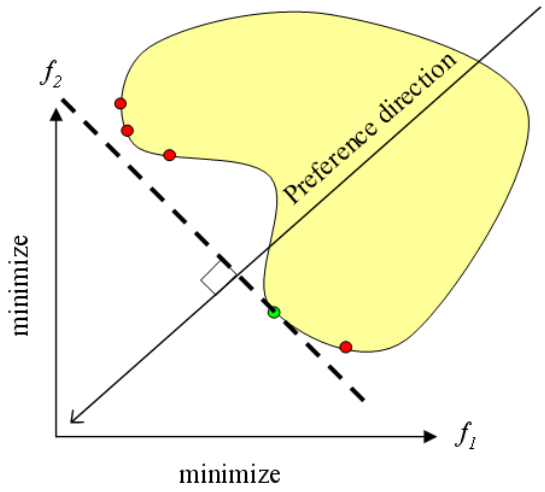


Figure 5-18: Example of the most preferred design selection for a two-dimensional minimization problem, both objectives weighted equally

For this research, the set of six $\vec{W}$ vectors shown in Table 5-8 are constructed to capture several DM preference structures. The sign for each weighting corresponds to

the optimization direction of the individual objective (positive for maximization and negative for minimization).

Table 5-8: Preference articulations

Number	$\vec{W}$	Description
1	$\left[\frac{1}{6} \quad \frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6}\right]$	best compromise
2	$\left[\frac{1}{2} \quad \frac{1}{2} \quad 0 \quad 0 \quad 0 \quad 0\right]$	best coverage
3	$\left[\frac{1}{3} \quad \frac{1}{3} \quad 0 \quad 0 \quad 0 \quad -\frac{1}{3}\right]$	best coverage, least risk
4	$\left[\frac{1}{4} \quad \frac{1}{4} \quad -\frac{1}{4} \quad 0 \quad 0 \quad -\frac{1}{4}\right]$	best coverage, least risk and time-of-flight
5	$\left[\frac{1}{5} \quad \frac{1}{5} \quad 0 \quad -\frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5}\right]$	best coverage, least risk and energy/propellant usage
6	$\left[\frac{1}{5} \quad \frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5} \quad 0\right]$	best coverage, least time-of-flight and energy/propellant usage

The first gives the best *compromise* design in the hypervolume by assigning equal weights to all objectives. The second vector assumes that the DM is only concerned with coverage performance at any cost and risk. The next vector assumes that coverage objectives and risk are all equally important. The fourth emphasizes time-of-flight equally with coverage and risk without regard for the propellant cost required to maneuver. The fifth, instead of minimizing time-of-flight for reconfiguration, weights propellant usage equally with coverage and risk. This vector produces the most preferred design without regard to the time-of-flight objective. The sixth and final preference structure weights coverage, time-of-flight, and propellant usage equally but does not

place a desire on reducing the number of satellites moved to achieve the new configuration. While these six do not represent all possible preference articulations, they do capture a balanced set that provide a context to a DM for the selection of a most preferred design.

The application of the preference structures of Table 5-8 to the 165 non-dominated solutions of Fig. 5-16 results in the most preferred designs shown in Table 5-9 for Case-2c-ideal.

Table 5-9: Most preferred designs, Case-2c-ideal (one plane lost, PDOP<6)

Preference Vector	Best Design	$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (m/s)	$f_5(\vec{x})$ (m/s)	$f_6(\vec{x})$
$\left[\frac{1}{6}\ \frac{1}{6}\ -\frac{1}{6}\ -\frac{1}{6}\ -\frac{1}{6}\ -\frac{1}{6}\right]$	41	1439.99	1439.03	3.56	48.79	0.00	1
$\left[\frac{1}{2}\ \frac{1}{2}\ 0\ 0\ 0\ 0\right]$	11	1439.99	1439.64	20.45	4.66	0.00*	4
$\left[\frac{1}{3}\ \frac{1}{3}\ 0\ 0\ 0\ -\frac{1}{3}\right]$	126	1439.99	1439.03	17.52	9.91	0.00	1
$\left[\frac{1}{4}\ \frac{1}{4}\ -\frac{1}{4}\ 0\ 0\ -\frac{1}{4}\right]$	79	1439.99	1439.03	1.06	163.10	0.00	1
$\left[\frac{1}{5}\ \frac{1}{5}\ 0\ -\frac{1}{5}\ -\frac{1}{5}\ -\frac{1}{5}\right]$	126	1439.99	1439.03	17.52	9.91	0.00	1
$\left[\frac{1}{5}\ \frac{1}{5}\ -\frac{1}{5}\ -\frac{1}{5}\ -\frac{1}{5}\ 0\right]$	125	1439.99	1439.41	3.51	28.41	0.00*	3

*The variance is not zero, but appears to be because of rounding to the nearest hundredth

Recall that the unlimited energy assumption was made in this section, so a reasonable first step towards selecting a final reconfiguration is to eliminate the most preferred designs that are clearly infeasible due to their excessive energy requirements. Using Eq. 4.9 and assuming a full maneuvering propellant budget of 25.0 kg, a dry mass of 1106 kg

and an engine with an impulse of 240 seconds, the maximum ΔV one could hope for is 52.6 m/s. Table 5-9 reveals that four of the preference structures result in identical coverage performance, and two of the four (Designs 79 and 41) exceed or nearly exceed the maximum possible ΔV. Additionally, the same most preferred solution (Design 126) appears for several preference vectors, which demonstrates that multiple preference structures can lead to a single design. The best coverage design, while taking nearly three weeks to achieve, does so with a reasonable amount of energy and almost no variance, but rephases four satellites in the process representing the most risky prospect. The final most preferred solution (Design 125), at worst, uses over half of potentially available propellant and maneuvers three satellites to get within 10 seconds of the best coverage solution. If time-of-flight is not critical, the DM should select Design 126 because only one satellite is maneuvered with the final configuration producing only a 20-second difference in the WCP-DVT when compared to the more risky and expensive Design 125.

A 20-second increase in coverage performance is what the DM would gain at greater cost and risk if they chose Design 125 over 126. To determine if the additional performance is worth the cost, the receiver coverage contour plots for Designs 125 and 126 are constructed in Figs. 5-19 and 5-20**,** respectively. The contour plot for the 'do-nothing' solution is shown in Fig. 5-21 for reference. Recall that the receiver grid was constructed for the Northern hemisphere only and, as such, the contour plots are displayed accordingly.

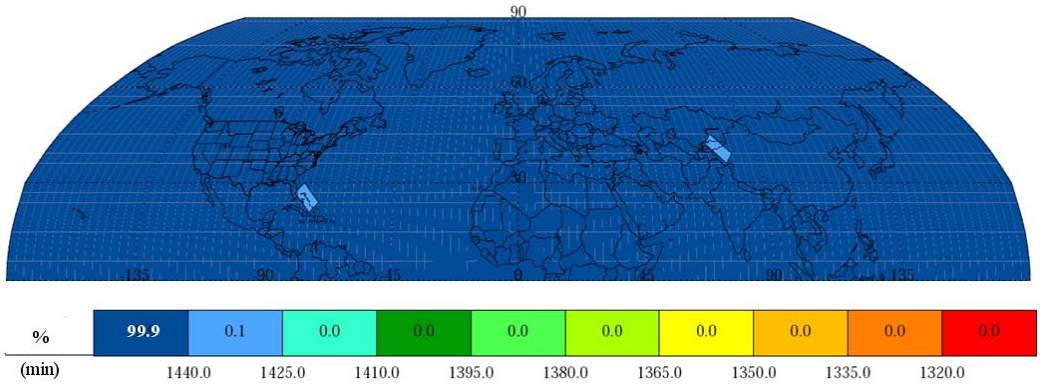


Figure 5-19: Case-2c-ideal (one plane lost, PDOP<6) global DVT contours for Design 125

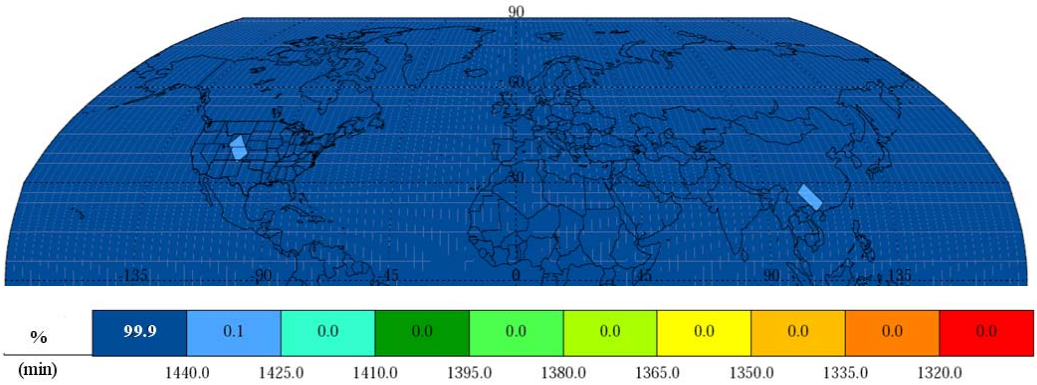


Figure 5-20: Case-2c-ideal (one plane lost, PDOP<6) global DVT contours for Design 126

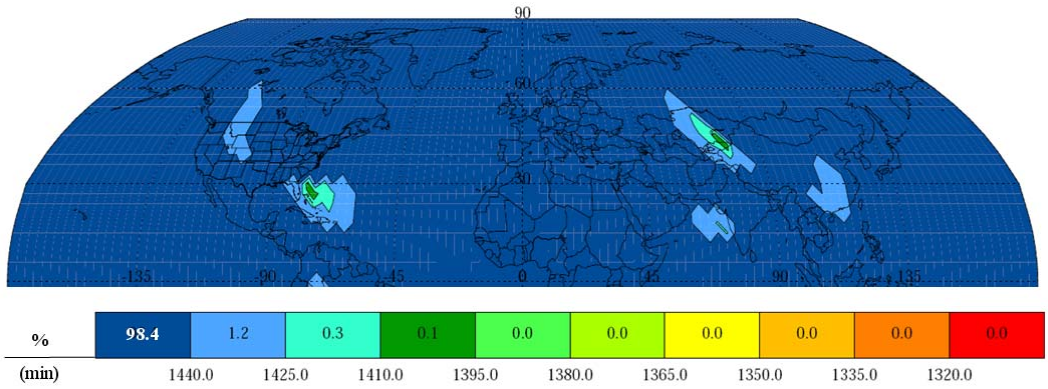


Figure 5-21: Case-2c-ideal (one plane lost, PDOP<6) global DVT 'do-nothing' solution

Both solutions provide significant improvement over the 'do-nothing' solution, and both have similar coverage gap patterns but over different regions of the hemisphere. If the DM is concerned with conterminous United States applications they should select design 125.

5.3.3.2 Access Constrained to PDOP ≤ 2

Case-2d-ideal, with the tightened PDOP constraint, finds 3636 solutions shown from two perspectives in Fig. 5-22 and 5-23 after searching for 406.1 hours and completing 15,089,931 function evaluations (18 interconnected runs). The range that the worst ΔV objective function takes has expanded several orders of magnitude with many designs requiring on the order of several km/sec in velocity changes.

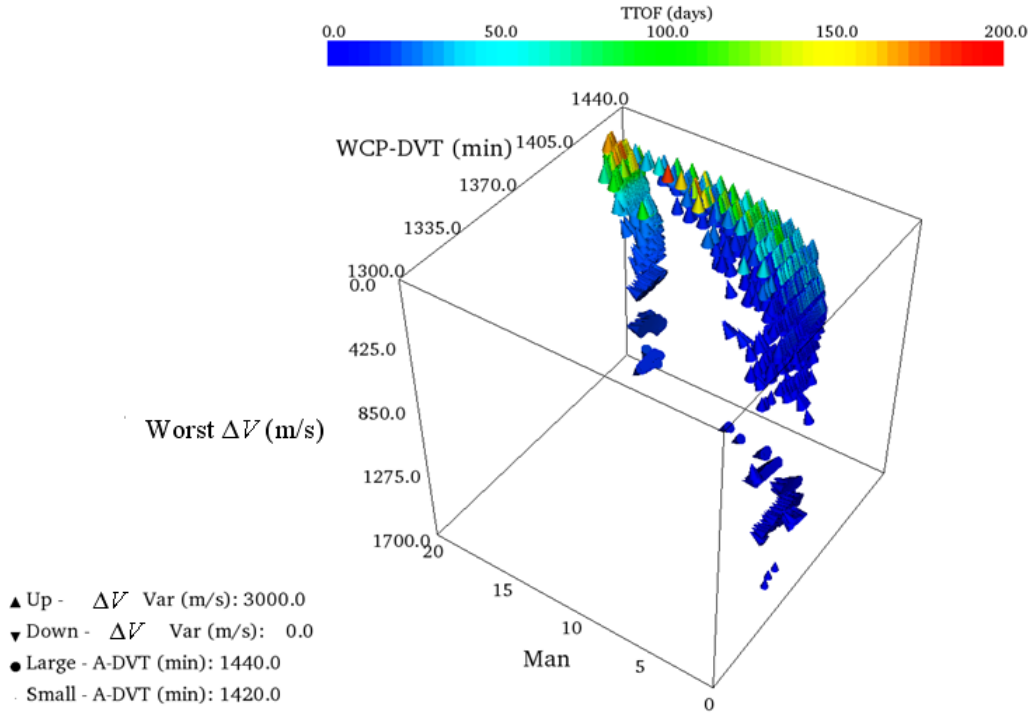


Figure 5-22: Glyph hypervolume at run termination, Case-2d-ideal (one plane lost, PDOP<2), view 1

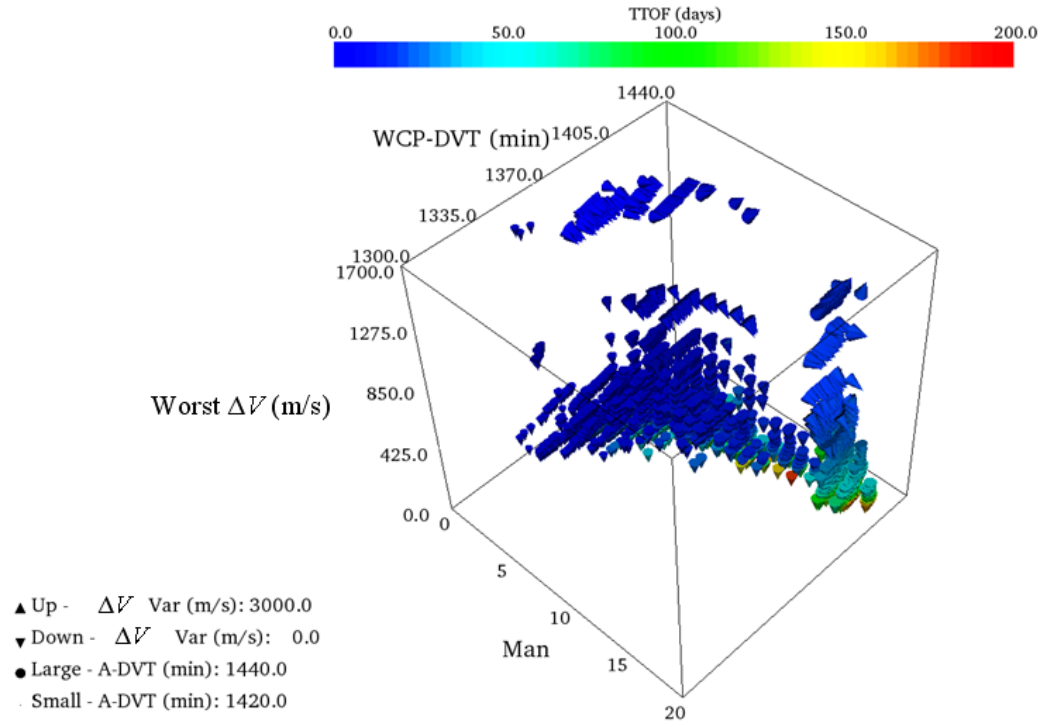


Figure 5-23: Glyph hypervolume at run termination, Case-2d-ideal (one plane lost, PDOP<2), view 2

Clearly, most of these designs would be eliminated in practice; however, it is instructive to observe the decision-variable plot shown in Fig. 5-24, where patterns begin to emerge. When maneuvering fewer than seven satellites, generally only those in adjacent planes to the one that was lost participate in the reconfiguration. As the number of satellites maneuvered approaches 19, designs in plane 2 contribute the most to extending the TTOF objective. This plot gives insight into the fitness landscape as well. There are no smooth transitions or precise patterns, but the decision variable space does exhibit some, albeit discontinuous, structure that the LC-ε-NSGA-2 is able to exploit.

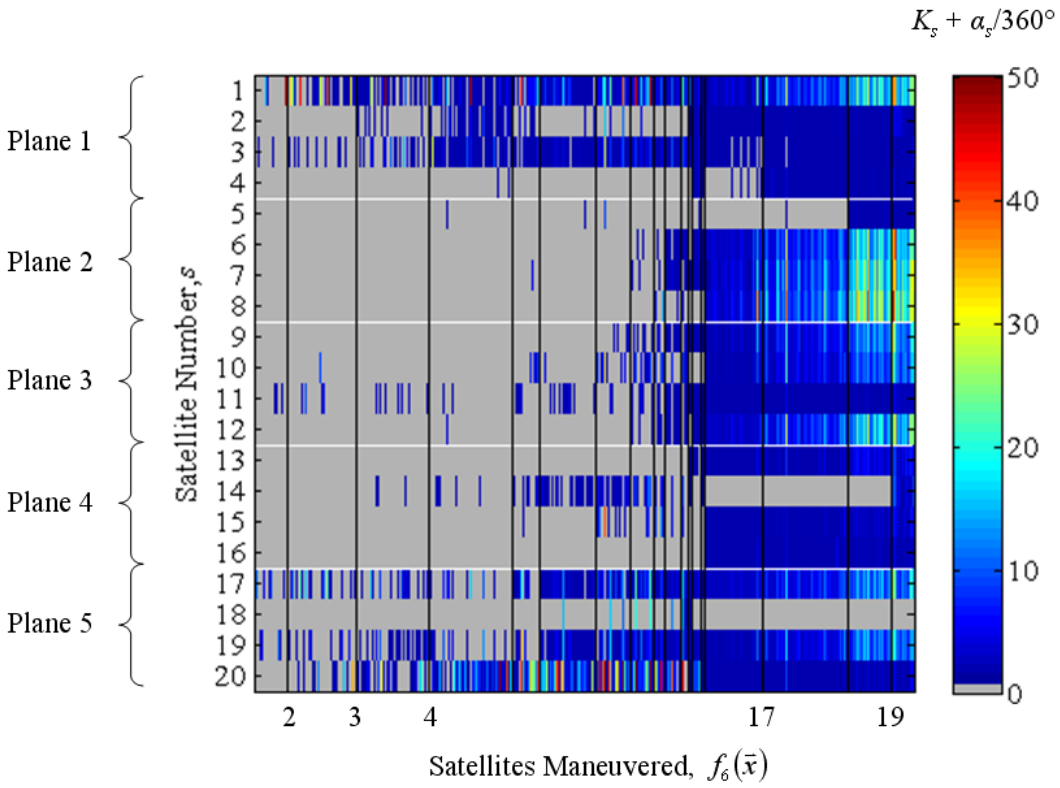


Figure 5-24: Hypervolume decision variable plot, Case-2d-ideal (one plane lost, PDOP<2) independent phasing orbits, K_s, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\vec{x})$ ascending

Applying the previously defined set of preference vectors to the hypervolume produces the set of most preferred designs shown in Table 5-10. Not surprisingly, all of the most preferred non-dominated solutions except Design 3395 require unrealizable amounts of energy. Even Design 3395 uses over 85% of the maximum potentially available and, because the variance is low, all six of the satellites maneuvered would use nearly the same 85% of their propellant. It is unlikely that Design 3395 could be

achieved; however, a contour plot of the coverage is generated in Section 5.3.4.2 for purposes of comparison with the operational case detailed in the same Section.

Table 5-10: Most preferred designs, Case-2d-ideal (one plane lost, PDOP<2)

Preference Vector	Best Design	$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (m/s)	$f_5(\vec{x})$ (m/s)	$f_6(\vec{x})$
$\left[\frac{1}{6}\ \frac{1}{6}\ -\frac{1}{6}\ -\frac{1}{6}\ -\frac{1}{6}\ -\frac{1}{6}\right]$	1169	1438.07	1411.17	14.59	110.08	5.00	6
$\left[\frac{1}{2}\ \frac{1}{2}\ 0\ 0\ 0\ 0\right]$	3247	1439.38	1424.81	50.24	223.00	66.59	18
$\left[\frac{1}{3}\ \frac{1}{3}\ 0\ 0\ 0\ -\frac{1}{3}\right]$	2259	1438.09	1411.19	7.61	376.00	83.85	6
$\left[\frac{1}{4}\ \frac{1}{4}\ -\frac{1}{4}\ 0\ 0\ -\frac{1}{4}\right]$	2244	1438.03	1411.11	5.62	1626.65	2030.91	6
$\left[\frac{1}{5}\ \frac{1}{5}\ 0\ -\frac{1}{5}\ -\frac{1}{5}\ -\frac{1}{5}\right]$	3395	1438.05	1411.03	36.53	43.42	0.790	6
$\left[\frac{1}{5}\ \frac{1}{5}\ -\frac{1}{5}\ -\frac{1}{5}\ -\frac{1}{5}\ 0\right]$	1210	1438.73	1419.16	18.21	76.08	5.58	14

Even though the initial application of the preference vectors resulted in mostly infeasible designs with regard to energy expenditure, there are many designs in the hypervolume that require ΔV less than 52.6 m/s. A data-thinning technique called brushing [109] is introduced here to eliminate any designs above a particular threshold in any of the objectives. Since energy is the issue, the brush is applied to $f_4(\vec{x})$ with a new ceiling set to 5 m/s. This value is chosen for illustration purposes only, but should be set according to the best knowledge of the propellant available to all satellites. The hypervolume is now reduced to 131 designs from the original 3395. Reapplying the preference vectors on the reduced set gives the set of solutions shown in Table 5-11.

Table 5-11: Most preferred designs, Case-2d-ideal after brushing to potential feasibility (one plane lost, PDOP<2)

Preference Vector	Best Design	$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (m/s)	$f_5(\vec{x})$ (m/s)	$f_6(\vec{x})$
$\left[\frac{1}{6} \quad \frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6}\right]$	1115	1434.53	1386.45	32.96	4.36	0.00*	2
$\left[\frac{1}{2} \quad \frac{1}{2} \quad 0 \quad 0 \quad 0 \quad 0\right]$	2612	1437.93	1411.10	165.97	4.91	0.01	12
$\left[\frac{1}{3} \quad \frac{1}{3} \quad 0 \quad 0 \quad 0 \quad -\frac{1}{3}\right]$	3182	1437.07	1401.64	96.23	4.77	0.01	6
$\left[\frac{1}{4} \quad \frac{1}{4} \quad -\frac{1}{4} \quad 0 \quad 0 \quad -\frac{1}{4}\right]$	1115	1434.53	1386.45	32.96	4.36	0.00*	2
$\left[\frac{1}{5} \quad \frac{1}{5} \quad 0 \quad -\frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5}\right]$	3182	1437.07	1401.64	96.23	4.77	0.01	6
$\left[\frac{1}{5} \quad \frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5} \quad 0\right]$	3521	1437.17	1402.53	77.79	4.91	0.00*	7

*The variance is not zero, but appears to be because of rounding to the nearest hundredth

Without exception, all designs in Table 5-11 have reduced performance in both coverage metrics when compared to the non-dominated solutions of Table 5-10. The best coverage design (2612) provides performance that is nearly as good as most of the solutions in Table 5-10, but is realized at the cost of reduction in the constellation's energy, increasing TTOF to a half a year, and doubling of the number of satellites maneuvered. In practice, the process of iterating on brushing levels and preference vectors until feasibility is obtained is cumbersome. The value of understanding the idealized cases, however, is that a performance ceiling is established because the hypervolumes represent the best attainable coverage without regard to energy limitations.

In the next section, the operational assumptions from Table 4-1 are imposed on the cases where one plane is lost, thereby eliminating the need for brushing of infeasible solutions.

5.3.4 One Plane Lost, Operational Assumptions

5.3.4.1 Access Constrained to PDOP ≤ 6

After searching to the end of the first checkpoint at 95.0 hours with 3,634,823 function evaluations (501 interconnected runs), Case-2c-op, finds the 55 solutions shown in Fig. 5-25. Note the axis label change from energy to kg of propellant. The decision variable plot, shown in Fig. 5-26, exhibits a more patterned behavior than the idealized case of Fig. 5-17. In the operational case, only four satellites (4, 6, 9, and 18) participate in any reconfiguration. Reviewing the initial propellant states for these satellites shows that they all had less than 10 kg of propellant of the maximum 25.0. Satellite 6 had the least propellant available of any of the satellites moved, at 4.3 kg, and also generally had the longest TTOF as illustrated by Fig. 5-26. When more than two satellites are maneuvered, satellite 6 no longer participates in reconfiguration; rather, satellites 4, 9, and 18 exclusively contribute to all three-vehicle solutions.

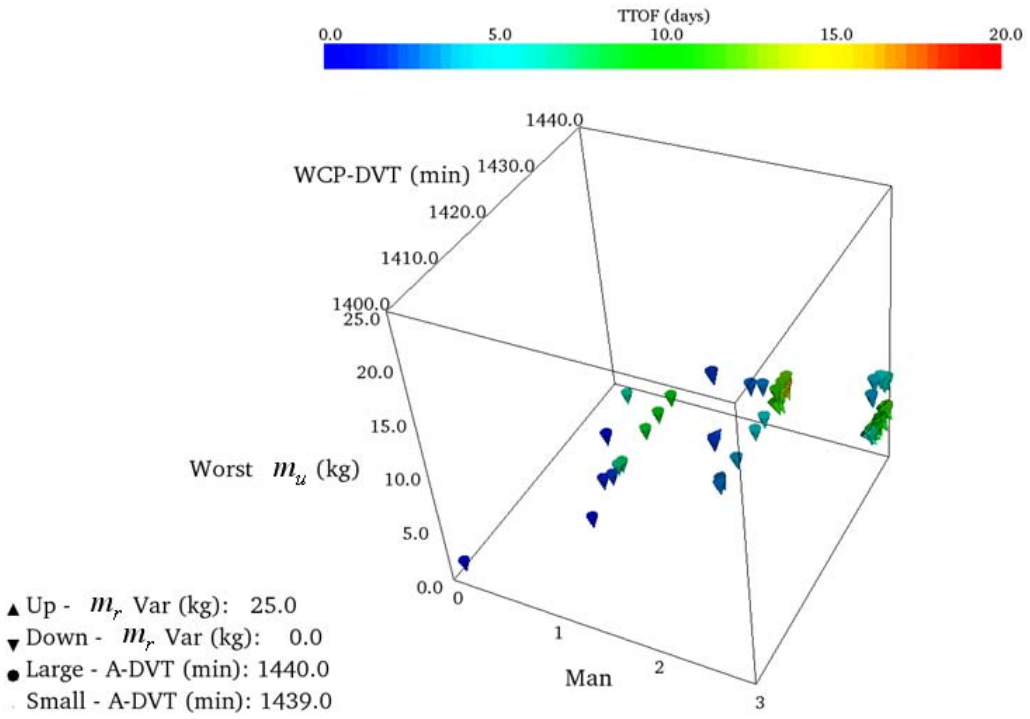


Figure 5-25: Glyph hypervolume at run termination, Case-2c-op (one plane lost, PDOP<6)

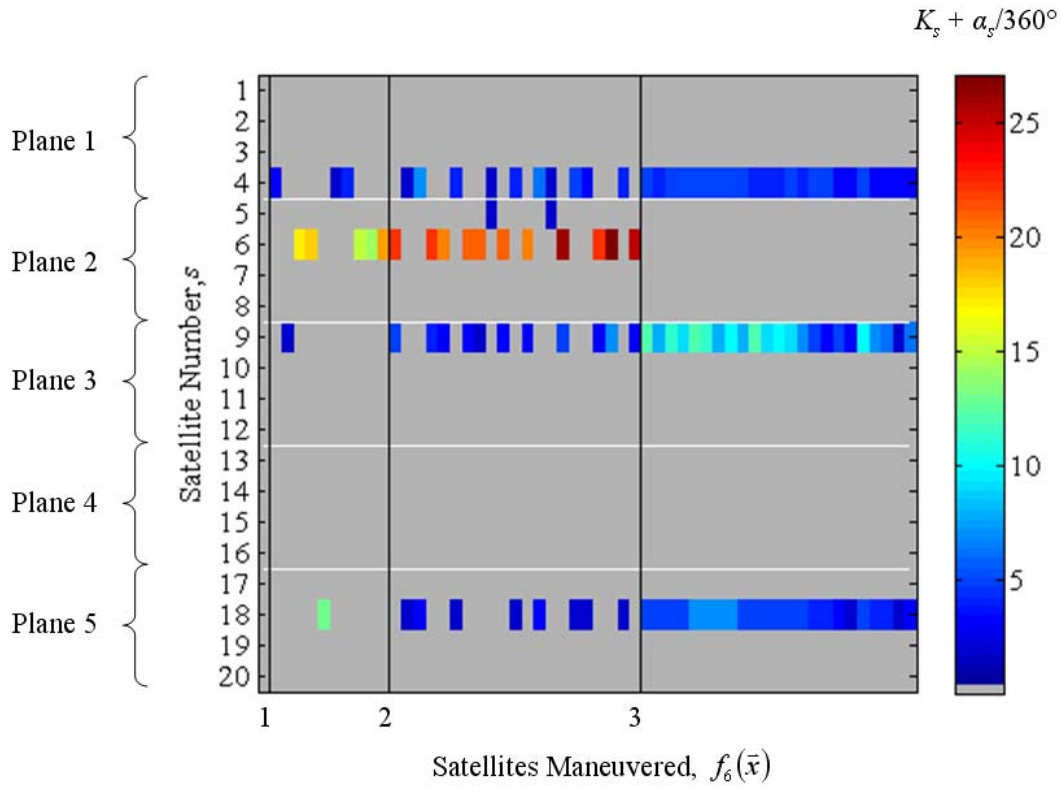


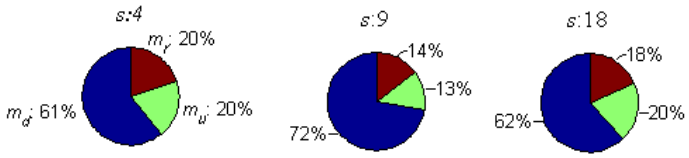
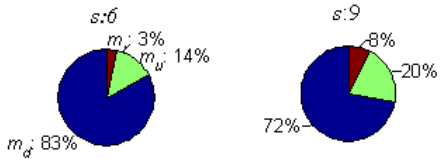
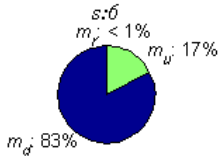
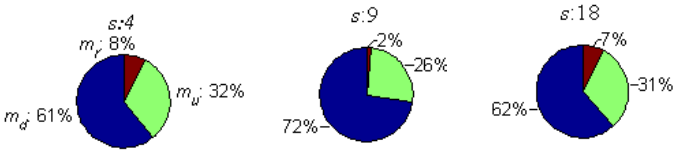
Figure 5-26: Hypervolume decision variable plot, Case-2c-op (one plane lost, PDOP<6) independent phasing orbits, K_s, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\bar{x})$ ascending

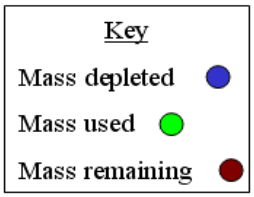
Applying the preference vectors to the hypervolume gives the solutions shown in Table 5-12. Interestingly, Design 30 appears three times indicating that the most preferred design, when coverage and risk are equally weighted, is invariant to preferences on TTOF or propellant costs. The remaining designs (15, 46, and 50) provide nearly continuous coverage but at degraded performance in the remaining objectives. In order to move towards a decision, the propellant-gage plots are generated for these designs in Table 5-13.

Table 5-12: Most preferred designs, Case-2c-op (one plane lost, PDOP<6)

Preference Vector	Best Design	$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (kg)	$f_5(\vec{x})$ (kg)	$f_6(\vec{x})$
$\left[\frac{1}{6} \quad \frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6}\right]$	30	1439.97	1432.51	9.51	4.28	0.0	1
$\left[\frac{1}{2} \quad \frac{1}{2} \quad 0 \quad 0 \quad 0 \quad 0\right]$	15	1439.99	1439.51	11.98	4.98	0.95	3
$\left[\frac{1}{3} \quad \frac{1}{3} \quad 0 \quad 0 \quad 0 \quad -\frac{1}{3}\right]$	46	1439.99	1439.03	17.00	4.96	0.61	2
$\left[\frac{1}{4} \quad \frac{1}{4} \quad -\frac{1}{4} \quad 0 \quad 0 \quad -\frac{1}{4}\right]$	30	1439.97	1432.51	9.51	4.28	0.0	1
$\left[\frac{1}{5} \quad \frac{1}{5} \quad 0 \quad -\frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5}\right]$	30	1439.97	1432.51	9.51	4.28	0.0	1
$\left[\frac{1}{5} \quad \frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5} \quad 0\right]$	50	1439.99	1439.23	7.00	7.89	0.47	3

Table 5-13: Propellant-gage plot (reduced to show only those maneuvered), hypervolume Designs 15, 46, 30 and 50, Case-2c-op (one plane lost, PDOP<6)

Design Number	Propellant Gages
15	
46	
30	
50	



Design 30 is not acceptable relative to the other choices since the maneuvering exhausts almost all of satellite 6 propellant available. Similarly, Designs 46 and 50 leave less than 10% propellant remaining. After considering the data shown in Table 5-13, the best choice is clearly Design 15, which also produces the best coverage performance of all the most preferred choices. Design 15, when compared to the idealized case with the same preference vector (Design 11 in Table 5-9) produces the same average performance and a worst-case-point coverage within 15 seconds of what is possible when the entire constellation has access to unlimited energy.

5.3.4.2 Access Constrained to PDOP ≤ 2

After 509.4 hours of search and 19,879,932 function evaluations (98 interconnected runs), 3401 solutions were found for Case-2d-op. The hypervolumes from two different perspectives are shown in Figs. 5-27 and 5-28 with the associated decision-variable plot given in Fig. 5-29. The hypervolumes reveal a clear correlation where, as the number of satellites maneuvered increases, so does the total time-of-flight and the variance of the mass of propellant remaining among those rephased.

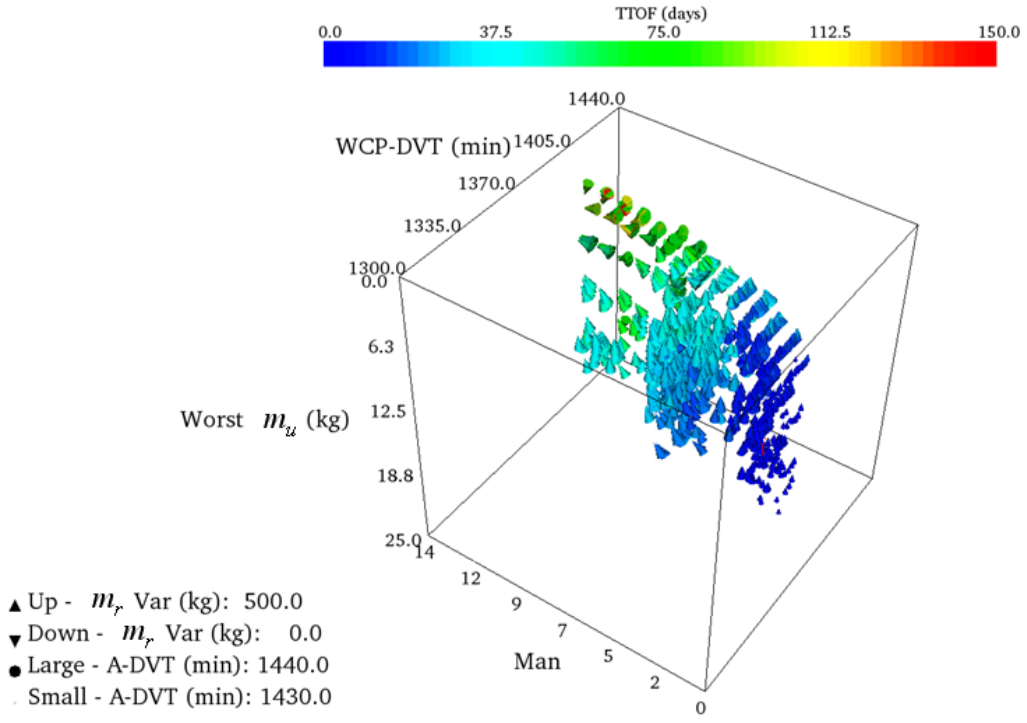


Figure 5-27: Glyph hypervolume at run termination, Case-2d-op (one plane lost, PDOP<2), view 1

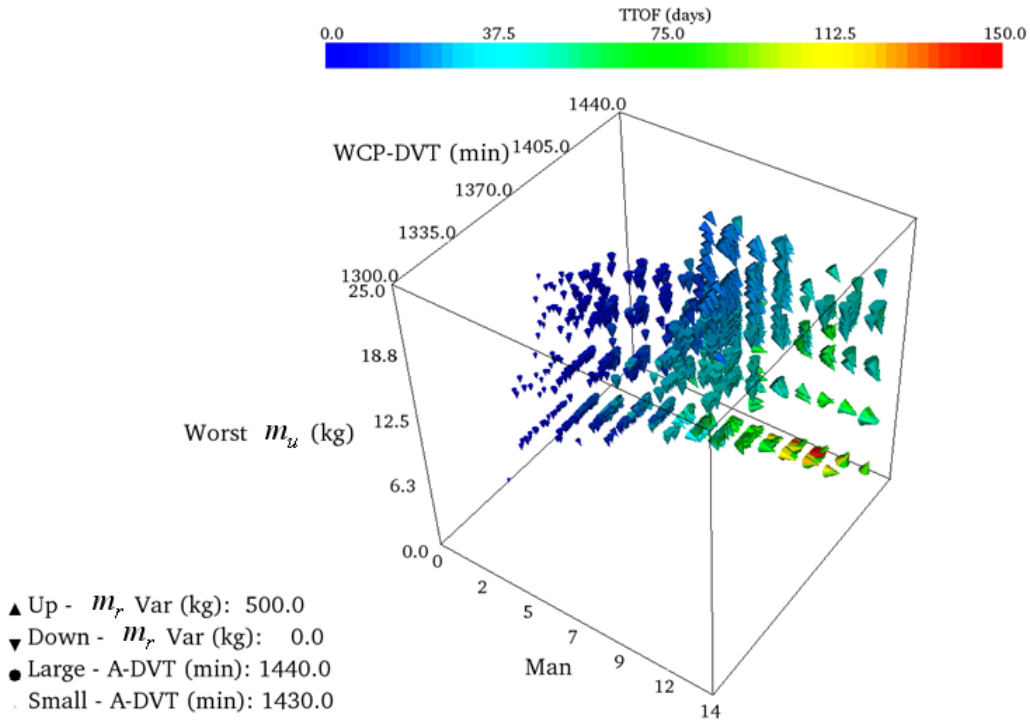


Figure 5-28: Glyph hypervolume at run termination, Case-2d-op (one plane lost, PDOP<2), view 2

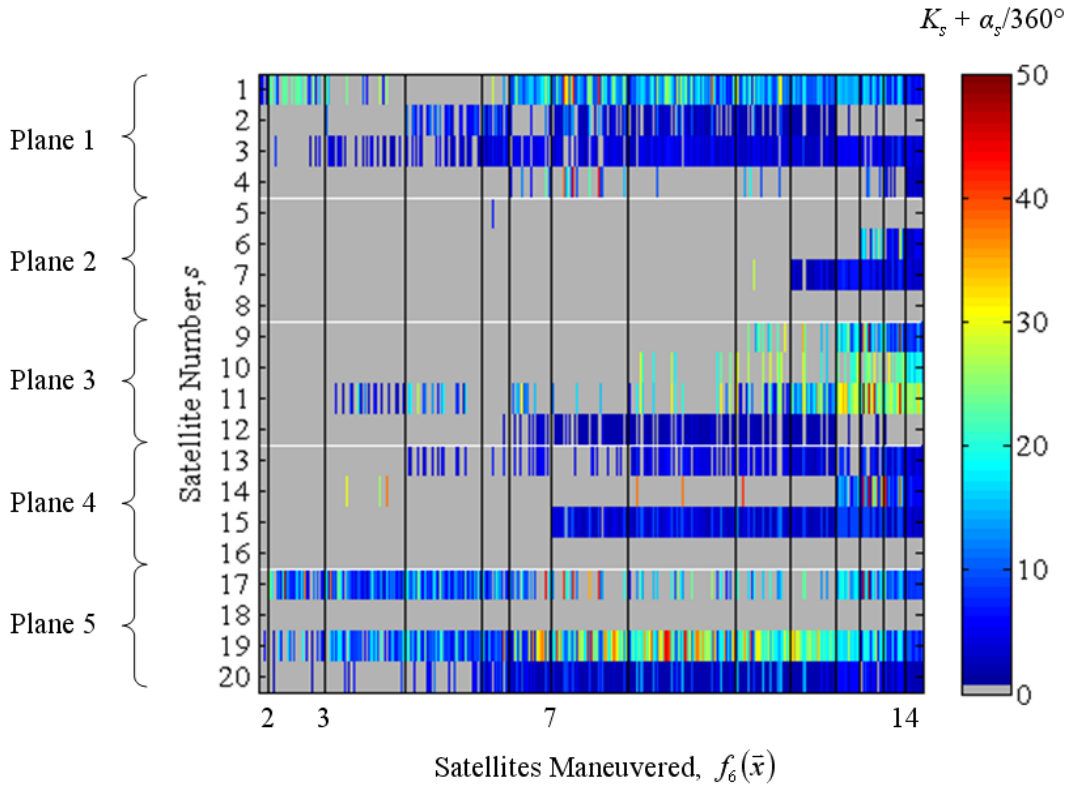


Figure 5-29: Hypervolume decision variable plot, Case-2d-op (one plane lost, PDOP<2) independent phasing orbits, K_s, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\bar{x})$ ascending

The key difference between the decision-variable plot of Fig. 5-29 and that with idealized assumptions (Fig. 5-24) is the fact that several satellites (8 and 16), are never maneuvered. Reviewing Table 4-1, satellites 8 and 16 have the least amount of propellant for maneuvering at 0.8 and 0.9 kg, respectively. This indicates that no benefit could be derived (in a non-dominated sense) from expending the relatively small quantity of propellant offered by these two vehicles. Satellite 18, on the other hand, has a reasonable amount of propellant at 9.5 kg, but is only maneuvered in two of the 3401 hypervolume solutions. This suggests that satellite 18, with the propellant available to it,

could generally not be rephased to benefit the constellation. Recalling Fig. 5-24, a similar observation is made where satellite 18 was almost never maneuvered, indicating that, even with unlimited energy, the same observation holds true. In Fig. 5-29, the maximum number of satellites maneuvered is 14, instead of all 19 in the unlimited energy case. Interestingly, all of these designs maneuvered the same 14 satellites and did not maneuver the same 6 (satellites 2, 5, 8, 12, 16, and 18). Satellites 2 and 5 had plenty of propellant available at 18.9 and 16.4 kg, respectively, but as before could not be rephased to produce a non-dominated result. This is an example of the non-intuitive discoveries that emerge through the application of the framework.

The preference vectors are applied to the hypervolume data and the resulting most preferred designs are shown in Table 5-14. This case exhibits the greatest difference in coverage performance among the set of most preferred designs relative to all other cases considered in this research.

Table 5-14: Most preferred designs, Case-2d-op (one plane lost, PDOP<2)

Preference Vector	Best Design	$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (kg)	$f_5(\vec{x})$ (kg)	$f_6(\vec{x})$
$\left[\frac{1}{6} \quad \frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6} \quad -\frac{1}{6}\right]$	2835	1436.10	1402.01	25.94	4.99	27.67	5
$\left[\frac{1}{2} \quad \frac{1}{2} \quad 0 \quad 0 \quad 0 \quad 0\right]$	2150	1438.38	1417.14	47.97	19.70	32.78	14
$\left[\frac{1}{3} \quad \frac{1}{3} \quad 0 \quad 0 \quad 0 \quad -\frac{1}{3}\right]$	3239	1438.07	1411.22	88.18	4.85	398.46	9
$\left[\frac{1}{4} \quad \frac{1}{4} \quad -\frac{1}{4} \quad 0 \quad 0 \quad -\frac{1}{4}\right]$	2039	1436.12	1399.18	12.55	13.74	29.87	4
$\left[\frac{1}{5} \quad \frac{1}{5} \quad 0 \quad -\frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5}\right]$	204	1436.26	1401.48	46.88	4.85	16.69	5
$\left[\frac{1}{5} \quad \frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5} \quad -\frac{1}{5} \quad 0\right]$	2849	1437.73	1408.17	39.71	9.37	14.15	8

The propellant-gage plots for each most preferred design are shown in Figs. 5-30–5-35. The best coverage design, shown in Fig. 5-31, requires that 14 satellites be maneuvered, during which time two (satellites 10 and 11) exhaust all but 1% of their useable propellant and another six (satellites 3, 4, 6, 14, 17, and 19) leave less than 10% in their tanks. It is unlikely that any DM would consider this design due to the dramatic reduction in life of the constellation at a relatively meager WCP-DVT coverage advantage over the next best most preferred design (3239). Designs 3239 and 2039, shown in Figs. 5-32 and 5-33, respectively, leave less than 1% of propellant remaining on satellites 10 or 11 and 17. Since these satellites might soon require de-orbiting and hence a reduction in capacity, they are discontinued from consideration. With all of the remaining most preferred designs, 2835, 204, and 2849, no satellites are maneuvered to the point of emptying the tanks.

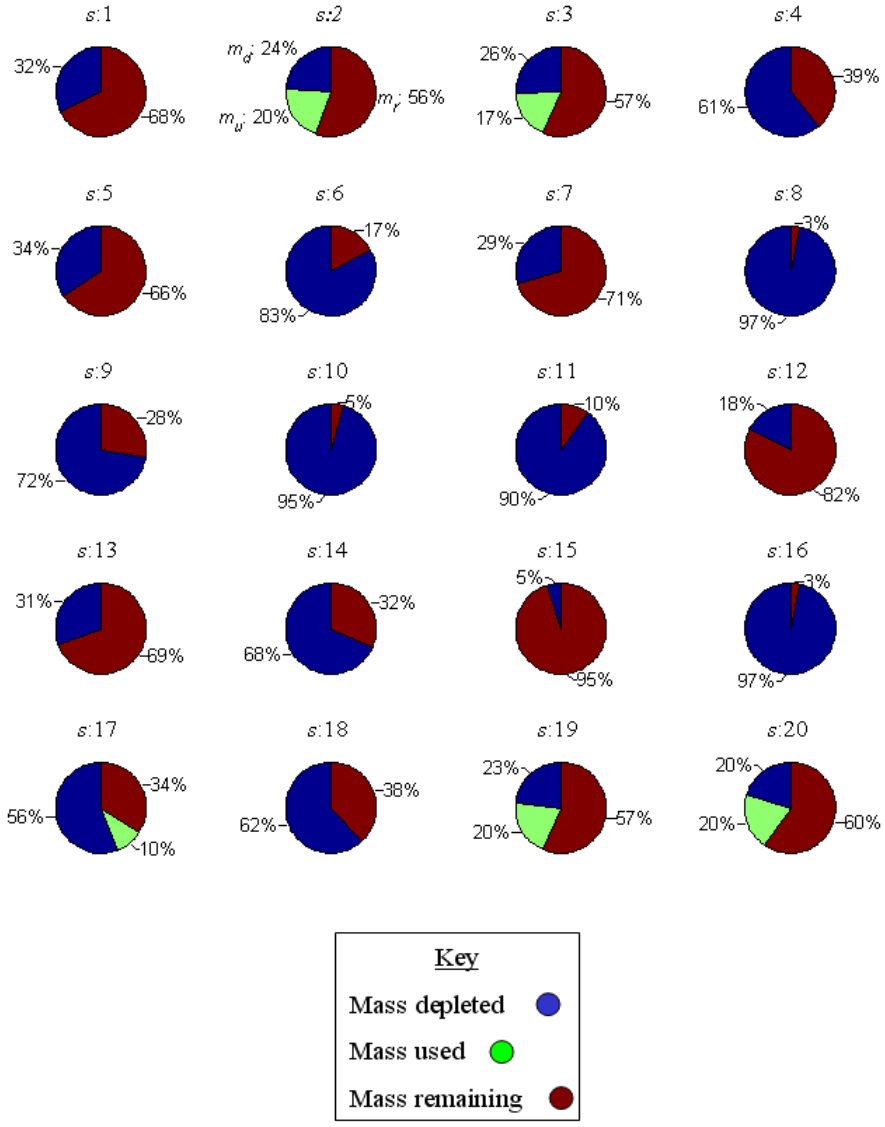


Figure 5-30: Propellant-gage plot, hypervolume Design 2835, Case-2d-op (one plane lost, PDOP<2)

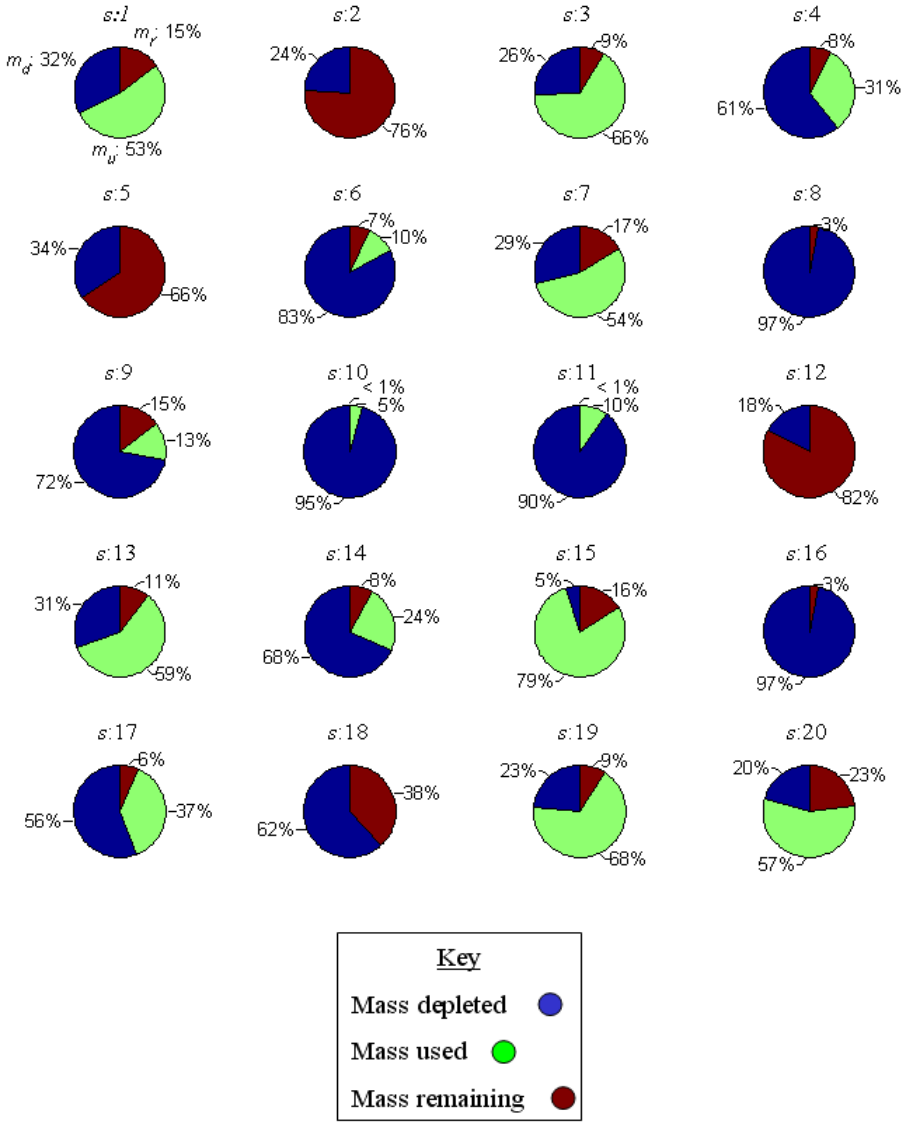


Figure 5-31: Propellant-gage plot, hypervolume Design 2150, Case-2d-op (one plane lost, PDOP<2)

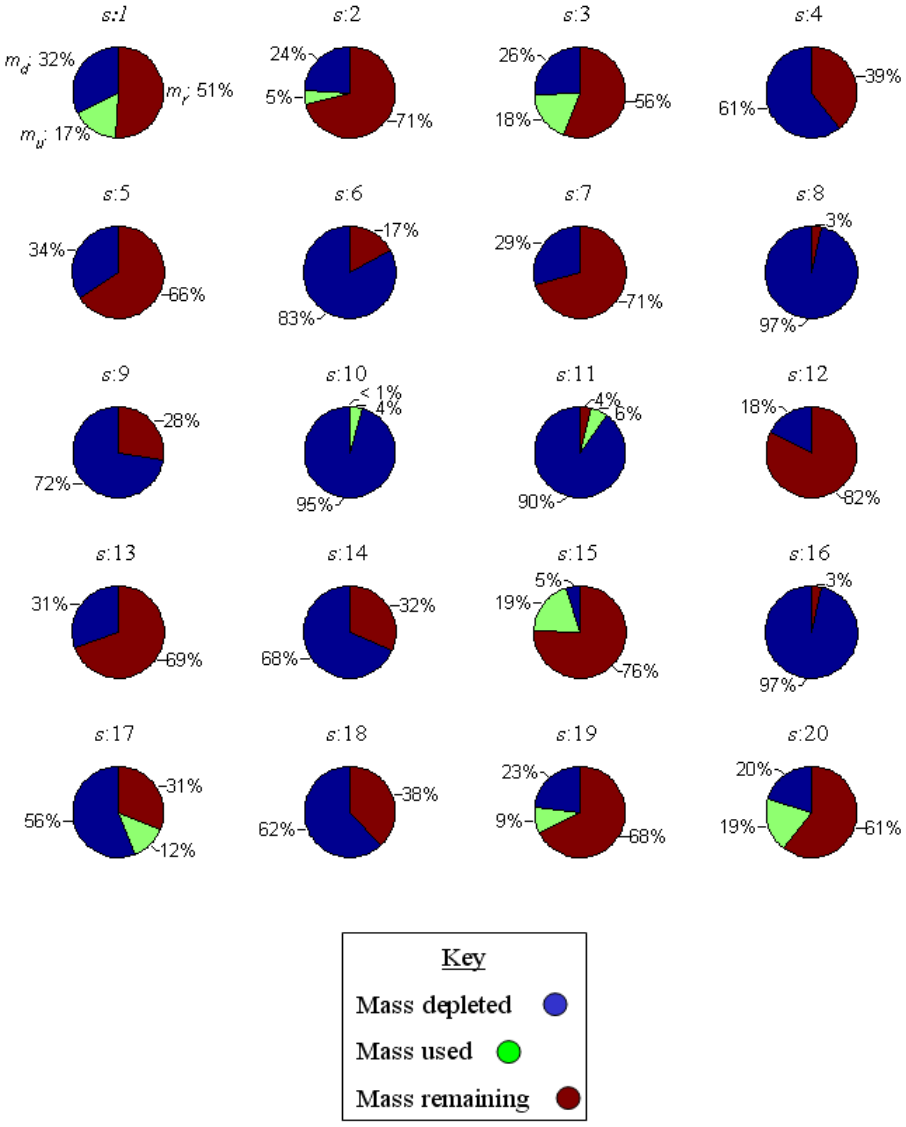


Figure 5-32: Propellant-gage plot, hypervolume Design 3239, Case-2d-op (one plane lost, PDOP<2)

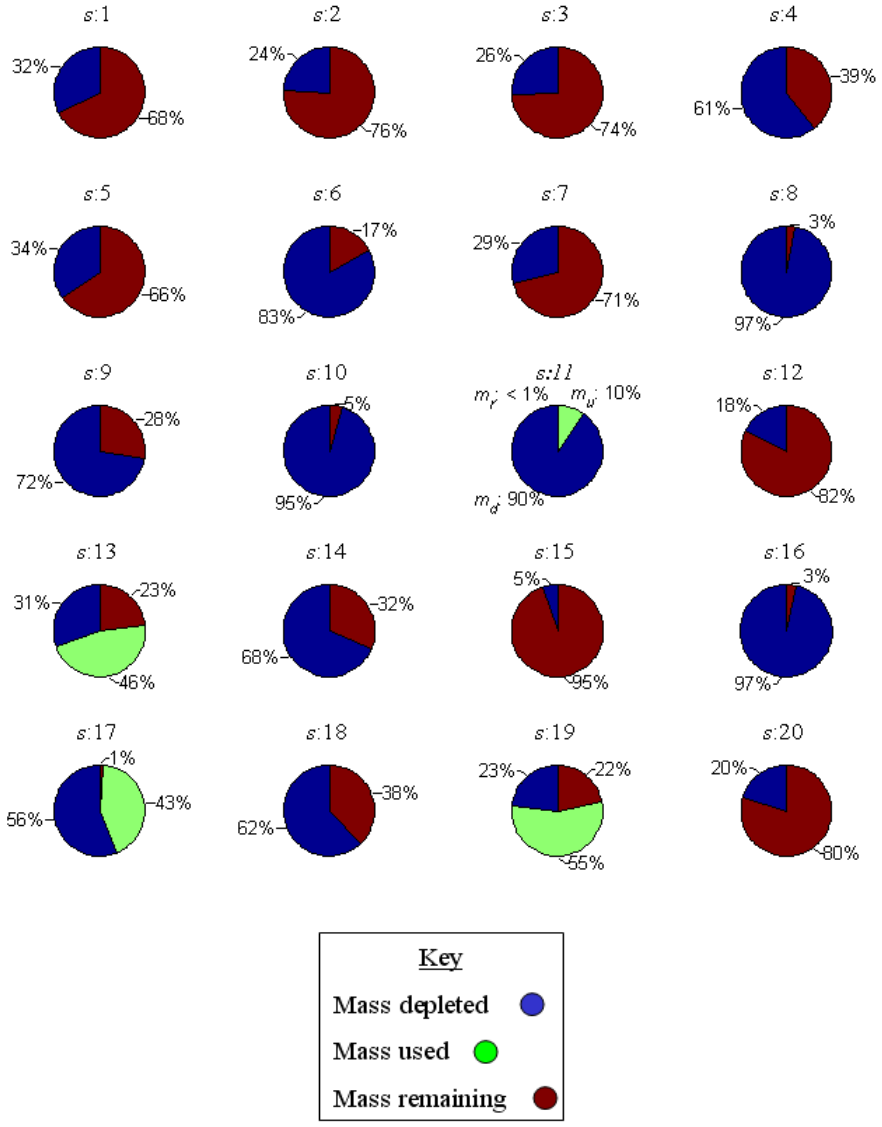


Figure 5-33: Propellant-gage plot, hypervolume Design 2039, Case-2d-op (one plane lost, PDOP<2)

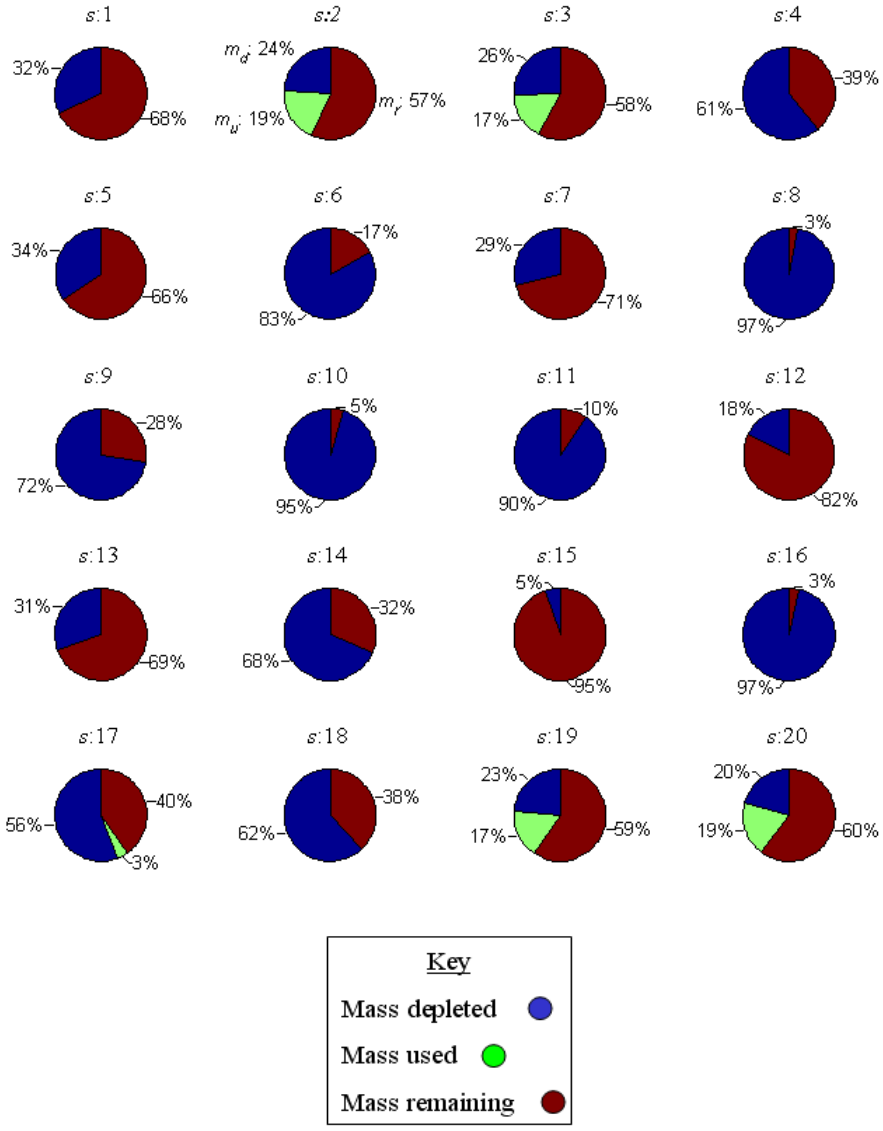


Figure 5-34: Propellant-gage plot, hypervolume Design 204, Case-2d-op (one plane lost, PDOP<2)

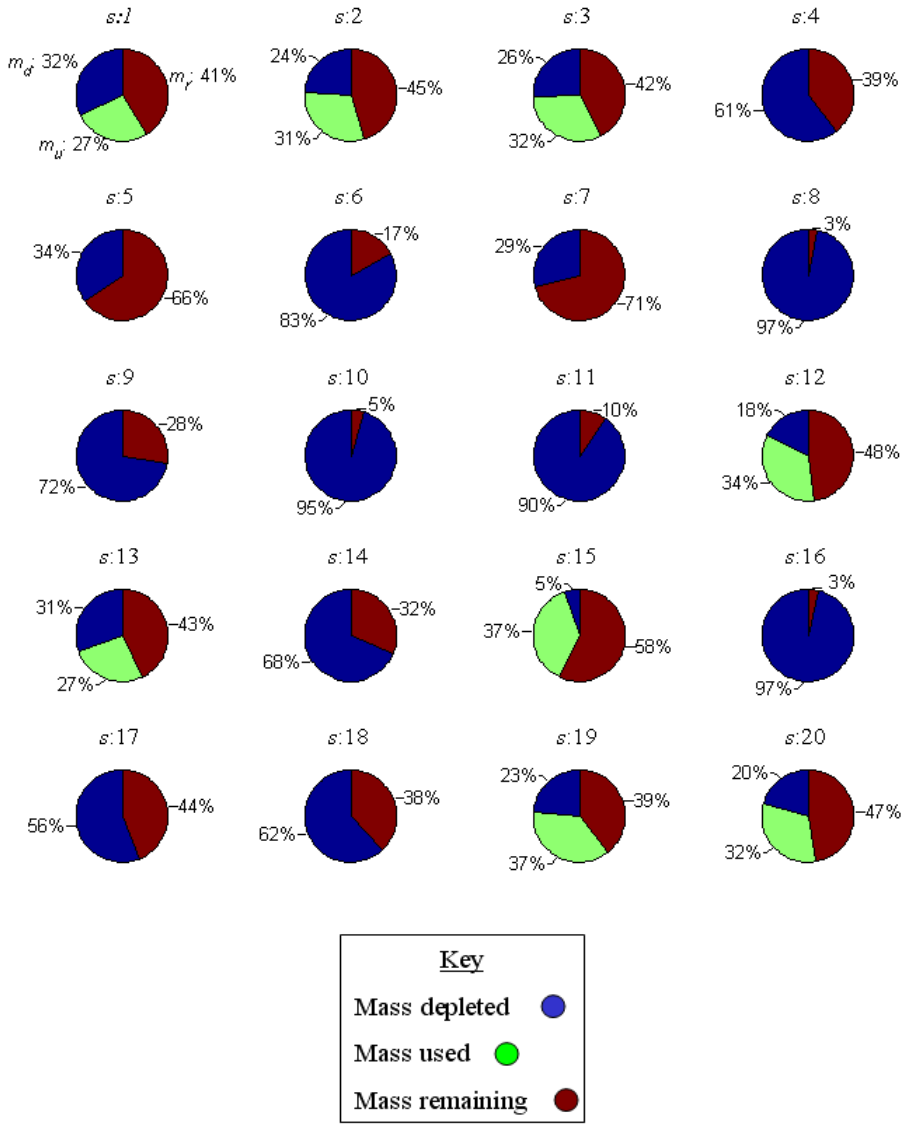


Figure 5-35: Propellant-gage plot, hypervolume Design 2849, Case-2d-op (one plane lost, PDOP<2)

In order to make a decision among the three remaining most preferred designs, refer again to Table 5-14. Design 2849 requires that nearly twice as many satellites be maneuvered as Designs 2835 and 204 and a TTOF of well over a month with nearly double the worst case propellant expenditure. What is gained is a 1.5-minute

improvement in A-DVT and about 7 minutes in WCP-DVT. If the DM selected this design, it would be the most costly of the three and they would need a compelling reason for the associated performance improvement. Notice that Designs 2835 and 204 have nearly identical coverage objective values and both maneuver the same five satellites, but Design 2835 completes the reconfiguration in half the time (about three weeks) at only a slightly greater propellant cost per satellite. At this stage in the decision making process, a DM would likely favor Design 2835, but before a final decision is made, it is prudent to consider the contour plots of all three solutions.

For comparison purposes, the contour plot of Design 3395 (Case-2d-ideal) is presented in Fig. 5-36 while Designs 2849, 204, and 2835 are shown in Figs. 5-37, 5-38, and 5-39, respectively. Recall, Design 3395 was the only *potentially* feasible design when unlimited energy was assumed. Design 2849 continuously covers 5% less of the receiver grid than the ideal case, while Designs 204 and 2835 degrade a further 5%, continuously covering approximately three-quarters of the grid. The coverage contours of Designs 204 and 2835 are almost identical. At this point, a DM has enough information to finalize their choice. Due to the high cost to achieve what amounts to a 6% gain in continuous coverage and the nearly identical contours of Figs. 5-38 and 5-39, Design 2835 is the best choice. Note, however, that Design 2849 provided significant coverage improvement to various regions like the conterminous United States. This might influence the DM to pay the cost. However, given this requirement, a better approach would be to modify the receiver grid to optimize the constellation towards only that region of interest.

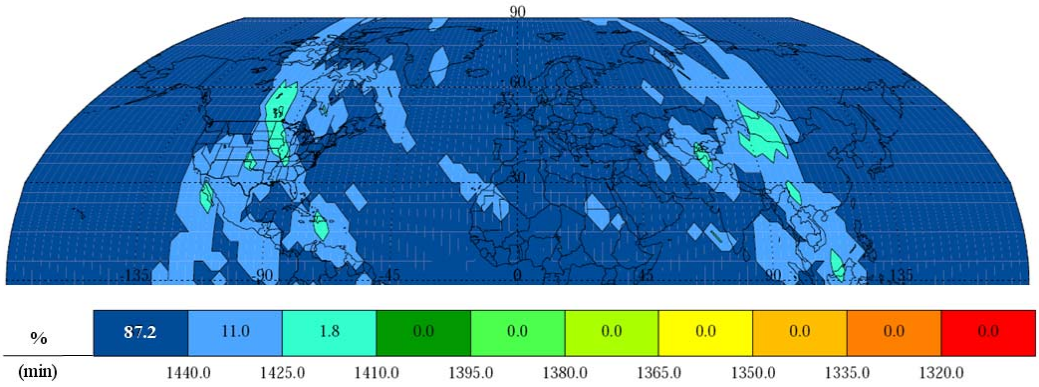


Figure 5-36: Case-2d-ideal (one plane lost, PDOP<2) global DVT contours for Design 3395

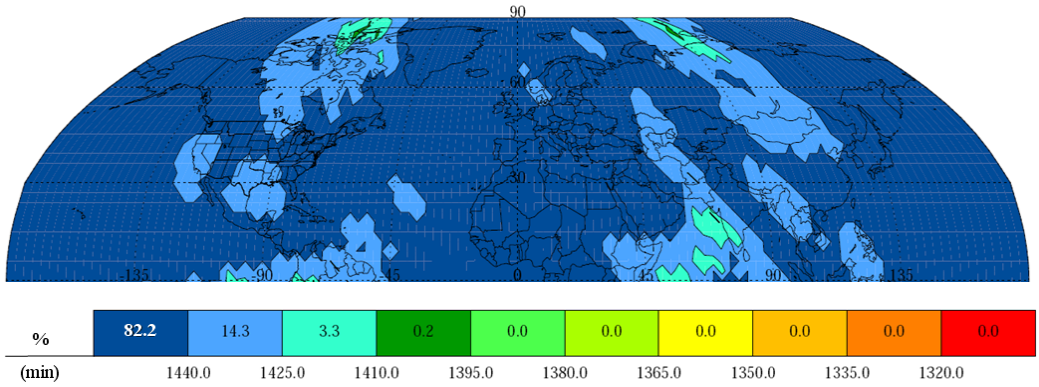


Figure 5-37: Case-2d-op (one plane lost, PDOP<2) global DVT contours for Design 2849

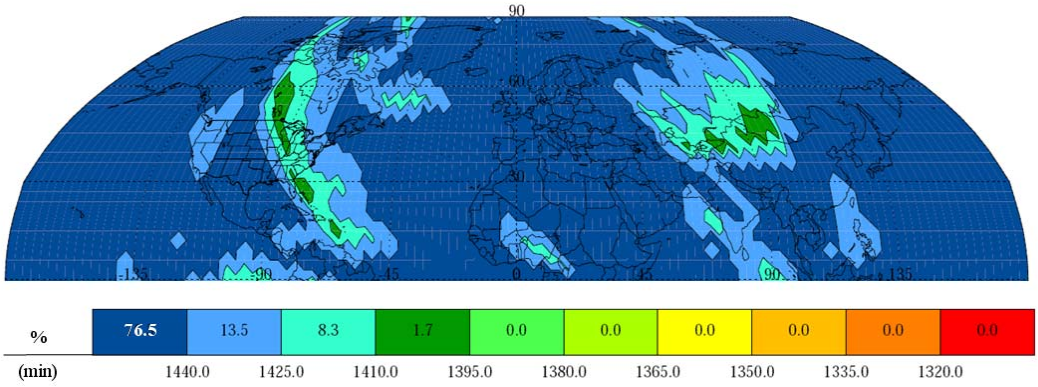


Figure 5-38: Case-2d-op (one plane lost, PDOP<2) global DVT contours for Design 204

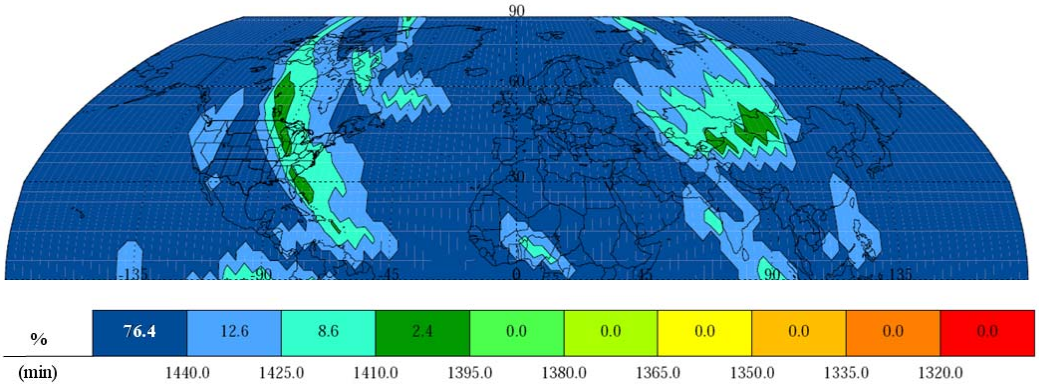


Figure 5-39: Case-2d-op (one plane lost, PDOP<2) global DVT contours for Design 2835

5.4 Revisiting Case-2d-op

Take a moment to review the search trace shown in Fig. 5-10 for Case-2d-op. As first mentioned in Section 5.2, this case represented an exception to the termination

criterion in that search progress was still being made even after over 500 hours of run time on 250 processing cores. If the termination of a run is used as a measure of the problem's difficulty, this case is the most challenging studied in this research. Case-2d-op is characteristic of a problem where the chosen epsilon vector resulted in extensive run time that is not likely acceptable in practice. One possible way to decrease the run time at the cost of fewer solutions available to a DM is to increase the coarseness of the epsilon vector.

Due to the inability to ascertain the geometrical properties of the six-dimensional fitness landscape, the coarse epsilon vector, shown in Table 5-15, is again formulated using engineering judgment, but also with some help from the previous optimization. In an iterative manner, various coarse epsilon vectors were used to post-process the 3401 designs of Section 5.3.4.2 until the set was reduced several orders of magnitude to 19 solutions. The final coarse epsilon vector selected for use in a new optimization simulation increases epsilon values in all objectives except $f_4(\vec{x})$ and $f_6(\vec{x})$. The epsilon value was decreased to 1 kg for $f_4(\vec{x})$ because a DM would likely favor more choices with regard to maximum propellant used on any maneuvered satellite. The epsilon value for $f_6(\vec{x})$ was left unchanged because again, in the author's judgment, this remains a dimension were more choices are desirable.

Table 5-15: Coarse epsilon vector comparison

Parameter	Setting
Original Epsilon vector	[1 min, 1 min, 1 hr, 5 kg, 1 kg, 1]
Coarse Epsilon vector	[2 min, 10 min, 30 days, 1 kg, 100 m/s, 1]

The optimization for Case-2d-op is repeated, but with the coarse epsilon vector (identified by 'coarse-eps'). The archive search dynamics for this new case are contrast against the original in Fig. 5-40.

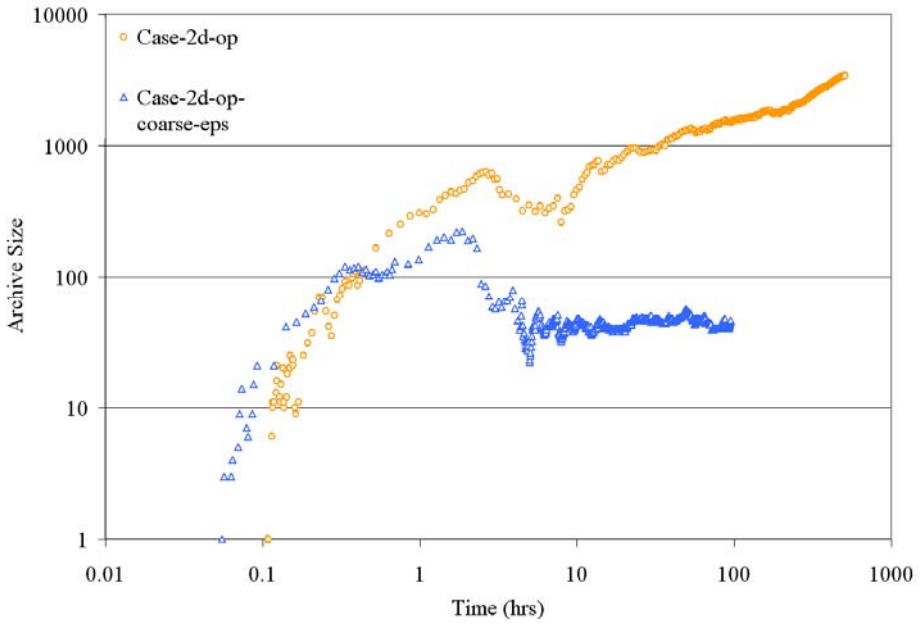


Figure 5-40: Archive dynamics, one plane lost, PDOP<2, *operational assumptions*, original and coarse epsilon vectors

After searching until the end of the first checkpoint at 95.0 hours and 3,716,083 function evaluations (951 interconnected runs), Case-2d-op-coarse-eps finished with 46 non-dominated designs. The archive size reaches an oscillatory steady state at approximately 40 hours into search. For any given problem, the epsilon vector uniquely fixes the maximum number of solutions that may be found (even if the knowledge of the fitness landscape is not known *a priori*). The oscillatory steady-state of the search dynamics trace provides an opportunity to determine if the population in the archive is mostly static

or being completely replaced between hours 40 and 95.0. Using hour 40 as the baseline, Table 5-16 shows both the number of original solutions and epsilon blocks remaining in the archive at 10 hour increments. In two dimensions, the epsilon values form a rectangle; in three, a volume is constructed; and greater than three dimensions results in hyper-blocks.

Table 5-16: Archive population replacement

Hour	Archive Size	Original solutions remaining	Original epsilon hyper-blocks remaining
40	47	47	47
50	54	10	38
60	47	5	28
70	44	3	25
80	40	2	21
90	41	2	21

Even though the archive is oscillating about a relatively fixed size, Table 5-16 shows that almost the entire population is replaced between hours 40 and 90. That is not the entire picture, however. Another perspective on search progress is provided by considering the epsilon hyper-blocks that remain as search proceeds.

Recall from Chapter 3 that each population member is assigned a box fitness and that no two solutions can exist within the same epsilon hyper-block. Table 5-16 shows that nearly half of the original hyper-blocks from the non-dominated population at hour 40 remain at hour 90. This data indicates that, while new non-dominated solutions are being found and replacing most of the archive, at least half of these improved designs still occupy identical epsilon hyper-blocks even after 50 hours of search. The fact that 21 of

the original 47 hyper-blocks remain also indicates that the algorithm is not moving to entirely different parts of the space, suggesting, at least empirically, that search progress has slowed. This observation provides confidence to terminate the run after the first checkpoint period of 95.0 hours has elapsed and compare the resulting hypervolume to Case-2d-op.

Figure 5-41 contrasts the hypervolumes resulting from the *a priori* application of (a) the coarse epsilon vector and (b) the original epsilon vector at run termination for both cases. At a glance, it appears that there are far fewer choices available to the DM due to the coarse epsilon vector, but before the observation can be confirmed, the quality of the solutions must be compared according to epsilon dominance. Applying the coarse epsilon vector of Table 5-15 *a posteriori* to the original hypervolume of Fig. 5-41b results in a reduction from 3401 solutions to 19 coarse epsilon non-dominated designs. When these 19 designs are combined with the 47 found for Case-2d-op-coarse-eps and sorted according to their coarse epsilon nondomination rank, 17 remain from (a) and 34 from (b). This result means that the original 500+ hour run for Case-2d-op had made significant search progress in many of the coarse epsilon hyper-blocks, but confirms the earlier suspicion that search had not reached a reasonable termination point. The 17 solutions remaining after coarse epsilon nondomination sorting of the original run replaced 13 of the 47 from Fig. 5-41a. This indicates that the extensive run time did result in better search progress in *certain regions* of the fitness landscape. In the coarse epsilon vector case, however, the 34 additional epsilon non-dominated designs shown in Fig. 5-41a suggests that the DM now has access to more diverse choice across a greater

expanse of the fitness landscape. This observation is highlighted more clearly by considering the decision variable plots for both cases.

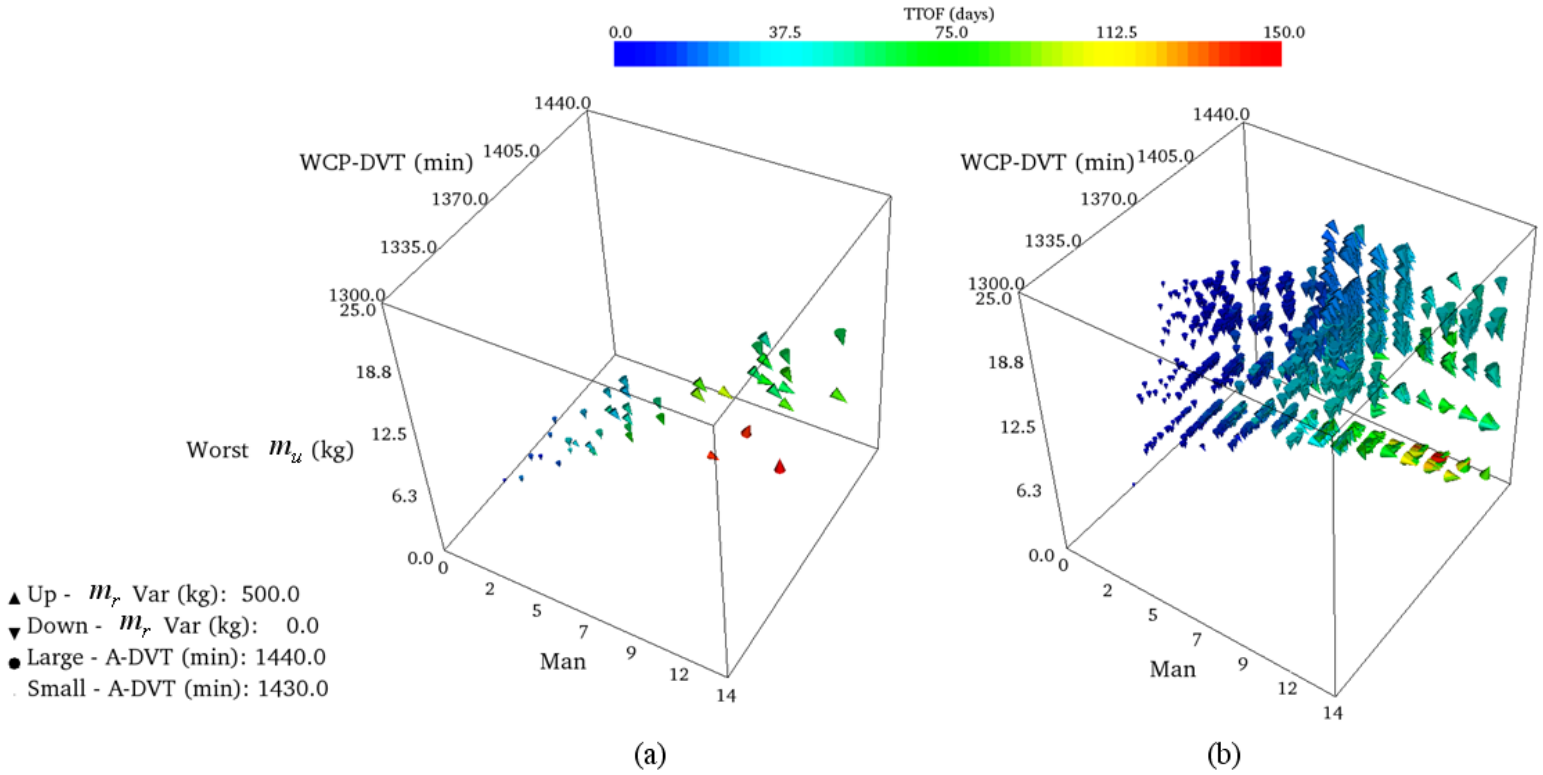
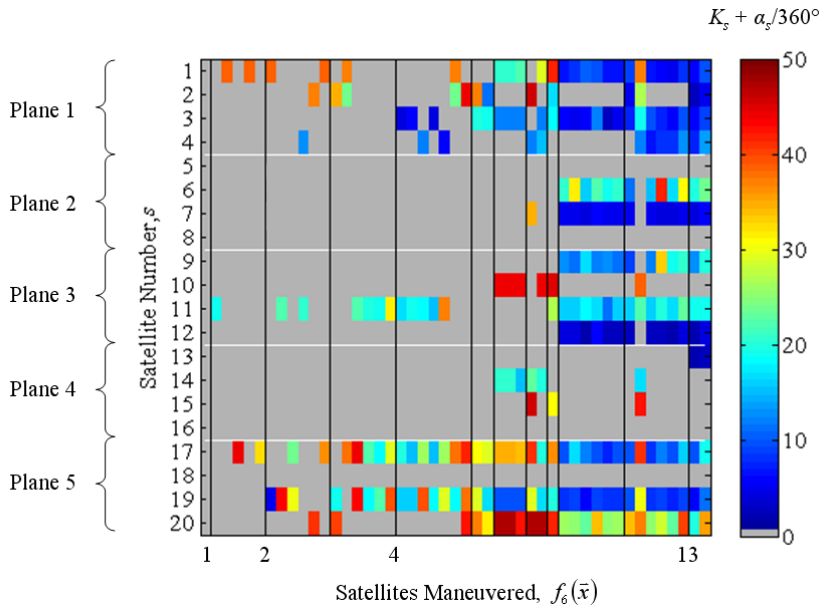
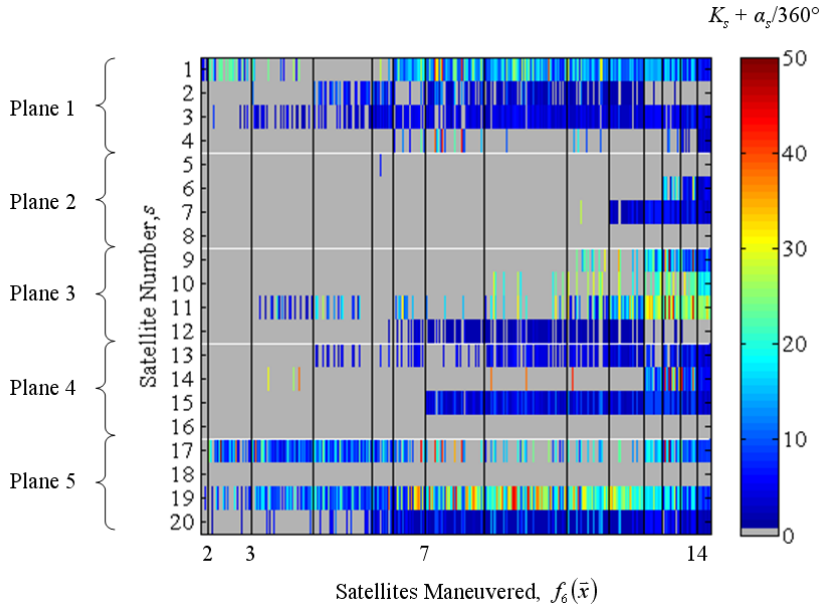


Figure 5-41: Glyph hypervolumes at run termination, Case-2d-op (one plane lost, PDOP<2), (a) coarse epsilon vector , (b) original epsilon vector

Figure 5-42 contrasts the decision-variable spaces produced by the final hypervolumes for both (a) coarse and (b) original epsilon vector cases. The hypervolume structure with regard to which satellites are maneuvered, and in what planes, is generally the same. However, over half of the solutions in Fig. 5-42a have TTOF greater than 50 days while only 20% of solutions in Fig. 5-42b fall into that bin. This observation is likely a direct consequence of the dramatic increase in the TTOF epsilon value from 1 minute to 30 days (43200 minutes). All of the other objectives receive increases or decreases in the epsilon vector that were several orders of magnitude less than that of TTOF. Another difference between Fig. 5-42a and 5-42b is that there are no solutions that maneuver more than 13 satellites with the coarse epsilon vector despite the fact that $f_6(\vec{x})$ epsilon value remained at 1 for both cases. Solutions in the fitness landscape that maneuvered 14 satellites were only found during the longer-duration, original epsilon vector search. At this point, it has been demonstrated that a coarse epsilon vector can allow for earlier run termination and provide more diverse choice across the fitness landscape, but no observations have yet been made regarding the specific objective function values.



(a)



(b)

Figure 5-42: Hypervolume decision variable plot, Case-2d-op (a) coarse epsilon vector, (b) original epsilon vector, (one plane lost, PDOP<2) independent phasing orbits, K_s, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\bar{x})$ ascending

When the preference structure, originally detailed in Table 5-8, is applied to the set of non-dominated designs in Fig. 5-41a, six solutions presented in Table 5-17 result. A design in Table 5-17 might be one of the 13 that were dominated by those of Case-2d-op. Keeping this in mind, it is instructive to compare Table 5-17 with 5-14 to illustrate the potential implications (if any) associated with choosing a coarse epsilon vector and shortening the search time.

Table 5-17: Most preferred designs, Case-2d-op-coarse-eps (one plane lost, PDOP<2)

Preference Vector	Best Design	$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (kg)	$f_5(\vec{x})$ (kg)	$f_6(\vec{x})$
$\left[\frac{1}{6}\ \frac{1}{6}\ -\frac{1}{6}\ -\frac{1}{6}\ -\frac{1}{6}\ -\frac{1}{6}\right]$	21	1434.01	1384.33	26.97	2.98	24.83	2
$\left[\frac{1}{2}\ \frac{1}{2}\ 0\ 0\ 0\ 0\right]$	32	1438.78	1420.20	89.52	8.00	170.26	13
$\left[\frac{1}{3}\ \frac{1}{3}\ 0\ 0\ 0\ -\frac{1}{3}\right]$	1	1437.96	1410.30	94.83	5.98	140.22	7
$\left[\frac{1}{4}\ \frac{1}{4}\ -\frac{1}{4}\ 0\ 0\ -\frac{1}{4}\right]$	33	1436.50	1391.26	26.01	5.88	104.95	4
$\left[\frac{1}{5}\ \frac{1}{5}\ 0\ -\frac{1}{5}\ -\frac{1}{5}\ -\frac{1}{5}\right]$	40	1436.04	1402.21	76.78	1.91	49.76	5
$\left[\frac{1}{5}\ \frac{1}{5}\ -\frac{1}{5}\ -\frac{1}{5}\ -\frac{1}{5}\ 0\right]$	40	1436.04	1402.21	76.78	1.91	49.76	5

After coarse epsilon nondomination sorting of combined results of Tables 5-17 and 5-14, all of the solutions from Table 5-17 remain, and two designs (204 and 2849) from Table 5-14 are dominated. Recall that the preference structure associated with the best compromise, Design 2835, was the best choice from Case-2d-op. This choice remains non-dominated when compared to the best compromise, Design 21, in Table

5-17. Even though Design 21 performs worst with regard to coverage, the reconfiguration uses less propellant and rephases only 2 satellites in about the same amount of time. In order to gain further insight into the propellant benefit offered by Design 21, its associated propellant-gage plot is presented in Fig. 5-43.

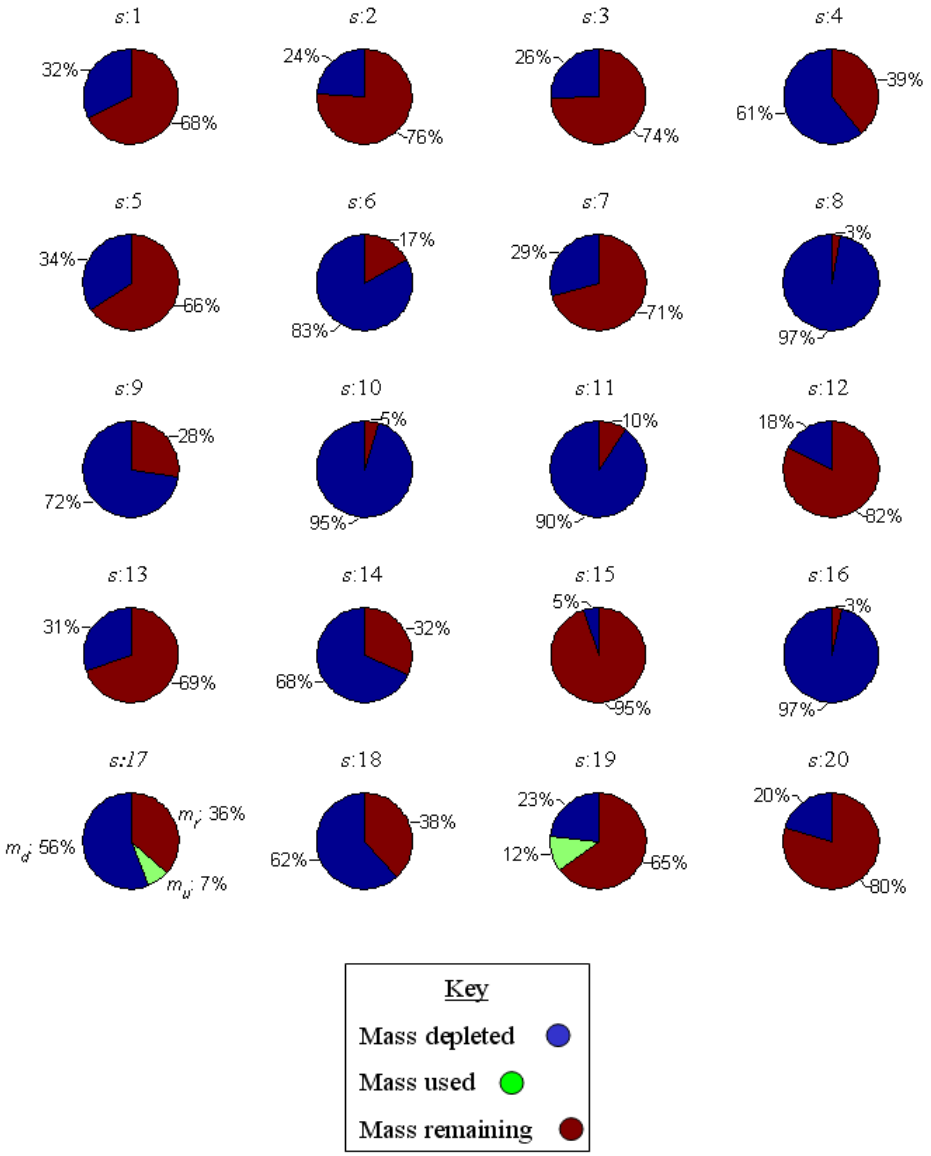


Figure 5-43: Propellant-gage plot, hypervolume Design 21, Case-2d-op-coarse-eps (one satellite lost, PDOP<2)

Clearly Design 21, which maneuvers fewer satellites and uses less propellant, provides an excellent alternative to Solution 2835. As before, the coverage contours are constructed in Fig. 5-44 for Design 21 to determine if the reduced performance relative to

Solution 2835 is worth the significant propellant savings and risk reduction. Comparing Fig. 5-44 with the contour of Solution 2835 in Fig. 5-39, coverage gaps do appear in areas that might be of importance to the decision maker (conterminous United States for example). However, there are regions, like northeastern Africa where coverage improves. There is no clear 'better' design here; however, the coarse epsilon vector did supply a DM with a design that would be a fine alternative if the computational or run time limitations prohibited the long-duration run with the original epsilon vector.

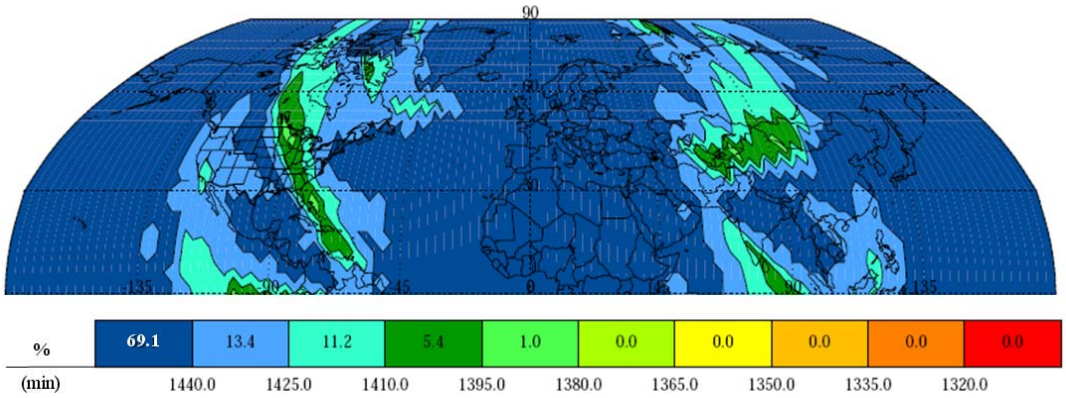


Figure 5-44: Case-2d-op-coarse-eps (one plane lost, PDOP<2) global DVT contours for Design 21

Several key lessons emerge from the observations and discussion presented in this section. A coarse epsilon vector can allow for run termination before that of a finer-grained case (taking care to study the box fitness changes over time to determine if search has slowed) and supply a greater diversity of solutions across the objective function landscape. Kollat and Reed [114] also demonstrated that a coarse epsilon vector can

result in compressed (in time) convergence. The shorter time-to-an-answer does not come for free, however. Seventeen of the 19 solutions remained from the long-duration run after being combined and sorted with the coarse hypervolume. This is an indication that certain parts of the fitness landscape were not yet explored and/or further exploration in regions with existing solutions would be beneficial. It appears that there is a tradeoff between providing an overview of the fitness landscape to the DM (coarse epsilon vector case) or providing better options within specific regions.

5.5 The Impact of a Global Grid

Smith [32] had noted, in reference to his research on GAs and satellite constellations, that the convergence time was directly tied to the number of variables to be optimized. In Section 5.2, it was shown that convergence time is not necessarily tied to the number of decision variables (in fact, it was demonstrated that increasing the number of variables can cause more rapid convergence). Another assumption that may extend function evaluation time and, therefore, convergence is the number of points in the receiver grid.

The contour plots of Section 5.1 demonstrated that PDOP constrained accesses produce coverage contours in the Northern hemisphere that are not symmetrical with the Southern hemisphere. This observation led to the conclusion that the typical hemispherical receiver grid assumed in the experimental design could not be used to *simulate* a global grid in order to reduce function evaluation time. As a result, all optimization runs presented to this point sought the reconfigurations that maximized

regional coverage performance to the Northern hemisphere. The peculiarity of the of the PDOP constraint effects on accesses presents an opportunity to study the impact that a *global* grid might have on extending run time and the corresponding reconfiguration alternatives.

Recall the equal area receiver grid defined in the experimental design consisted of 2339 points. For this section, a *global* equal area grid is constructed where receivers are placed at 0º latitude spaced 3º degrees apart from 0º to 360º East longitude at 0 km altitude. The remaining receivers are placed over the globe such that an equal area grid is formed totaling $L = 4558$ points.

Case-1d-op, where one satellite was lost and PDOP constrained to 2 or better, is selected to study the impact of the global grid. The archive size dynamics for this case (Case-1d-op-globe) are represented in Fig. 5-45 along with the original search trace. Figure 5-45 reveals an unexpected result. Doubling of the receiver grid approximately doubled the evaluation time for each objective vector; however, the termination criterion is still met within the first 95.0-hour check point period. At run termination, Case-1d-op-globe completed 1.64 million function evaluations and contained ten solutions in the archive. The hemispherical receiver grid of Case-1d-op completed twice as many function evaluations (3.27 million) in that time and finished with five archive solutions.

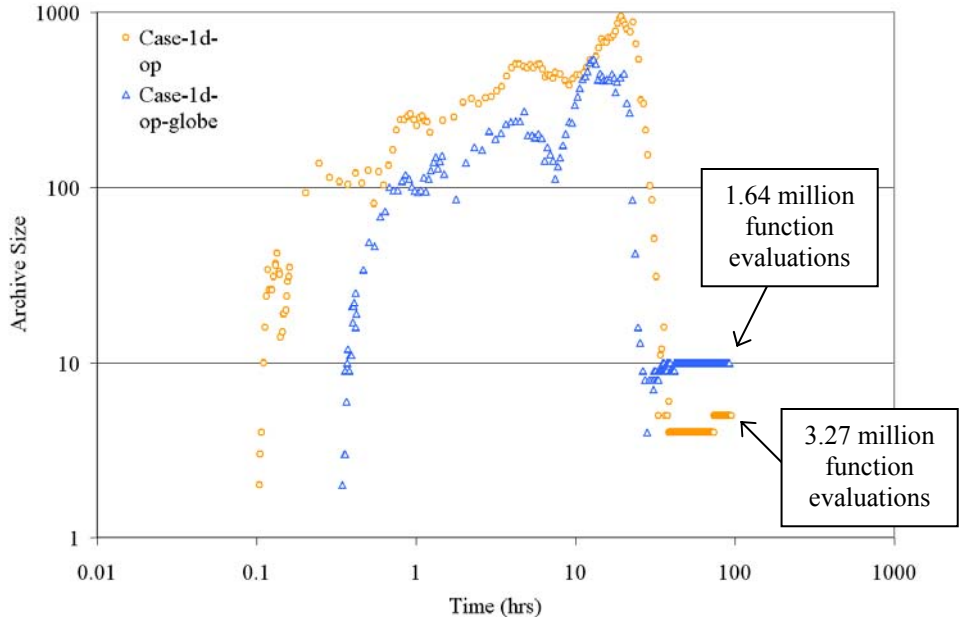


Figure 5-45: Archive dynamics, one plane lost, PDOP<2, *operational assumptions*, global receiver grid

This result indicates that the global receiver grid made the problem *easier* to solve, despite the increase in function evaluation time. Another interesting observation from Fig. 5-45 is that no feasible solutions were found for the global grid case until about one-half hour into search, which is nearly triple that of Case-1d-op. The key lesson here is that no correlation should be assumed between the sizes of a receiver grid and the difficulty or run time for the problem.

The objective vectors for the ten solutions making up the hypervolume at run termination in Fig. 5-46a are given in Table 5-18. The original hypervolume for Case-1d-op is supplied in Fig. 5-46b for comparison purposes.

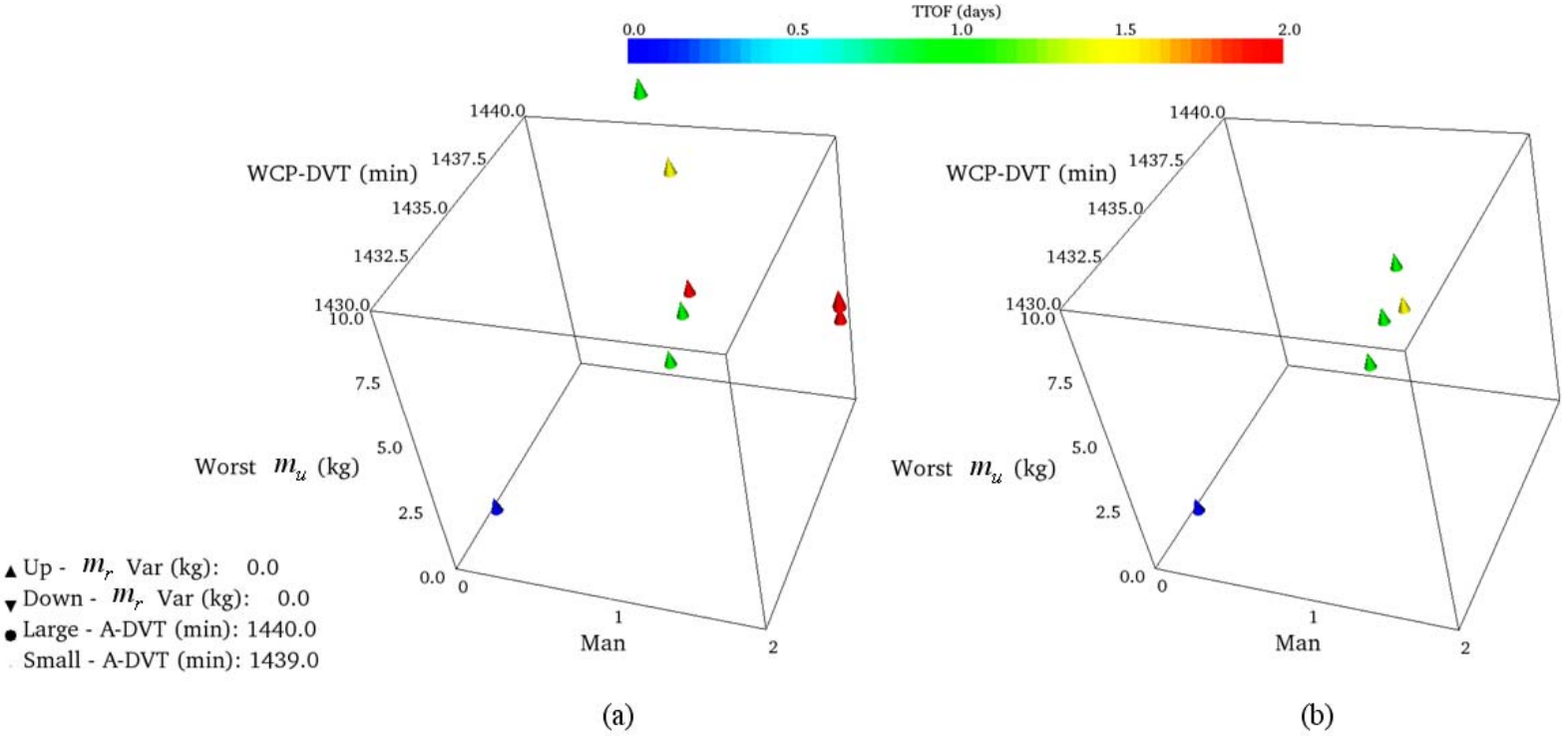


Figure 5-46: Glyph hypervolumes at run termination, Case-1d-op (one satellite lost, PDOP<2), (a) global receiver grid, (b) hemispherical receiver grid

Table 5-18: Hypervolume objective vectors, Case-1d-op global receiver grid (one satellite lost, PDOP<2)

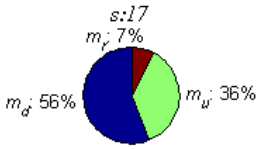
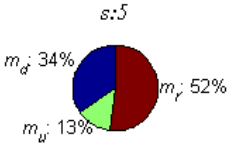
Design	$f_1(\vec{x})$ (min)	$f_2(\vec{x})$ (min)	$f_3(\vec{x})$ (days)	$f_4(\vec{x})$ (kg)	$f_5(\vec{x})$ (kg)	$f_6(\vec{x})$
1	1439.9997	1439.06	2.00	4.85	6.00	2
2	1439.9947	1437.00	1.00	3.46	0.00	1
3	1439.9998	1439.23	1.51	9.10	0.00	1
4	1439.9998	1439.23	3.00	4.57	0.00	1
5	1439.9997	1439.06	2.50	4.90	2.00	2
6	1439.9997	1439.06	2.00	4.26	12.28	2
7	1439.9882	1432.84	0.00	0.00	0.00	0
8	1439.9997	1439.06	3.00	4.77	1.00	2
9	1439.9983	1438.09	0.99	12.38	0.00	1
10	1439.9989	1438.46	1.00	4.26	0.00	1

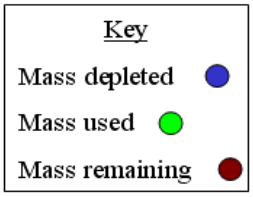
In Section 5.3.2, it was shown that two of the reconfiguration designs for Case-1d-op would result in continuous four-fold coverage to the entire hemispherical receiver grid. When a global grid is used, continuous coverage is no longer possible with any reconfiguration of the remaining satellites. Additionally, the hypervolume for the global case (Fig. 5-46a) includes four solutions that maneuver two satellites, whereas the previous case (Fig. 5-46b) rephased a maximum of one vehicle. These four solutions provide a WCP-DVT within one minute of each with various tradeoffs between propellant usage and time-of-flight. They are not, however, the best coverage solutions. Designs 3 and 4 maneuver one satellite and come within 40 seconds of continuous global coverage.

Table 5-19 compares the propellant-gages for designs 3 and 4 and the best coverage design from Table 5-7 for the hemispherical coverage case. Table 5-19 shows

that the same satellite (17) is maneuvered to achieve the best global coverage, with the only difference between the designs being a longer TTOF for Design 4. The DM must choose between reconfiguration time of 1.51 (Design 3) or 3.00 days (Design 4). It is likely that Design 4 would be the best choice since one-quarter of available propellant would remain after the maneuver.

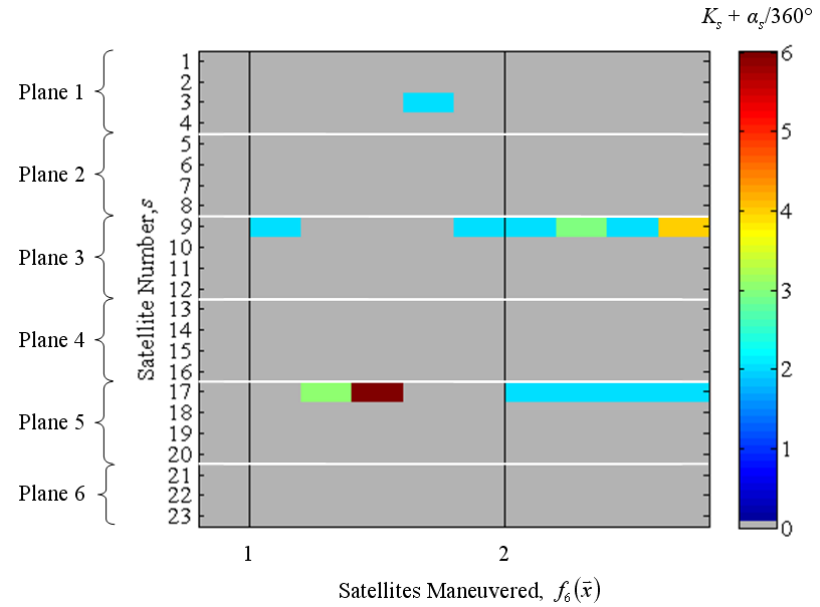
Table 5-19: Propellant-gage plot, hypervolume Designs 3, and 4, Case-1d-op-globe (one satellite lost, PDOP<2)

Design	Propellant-gage
3	
4	
2*	

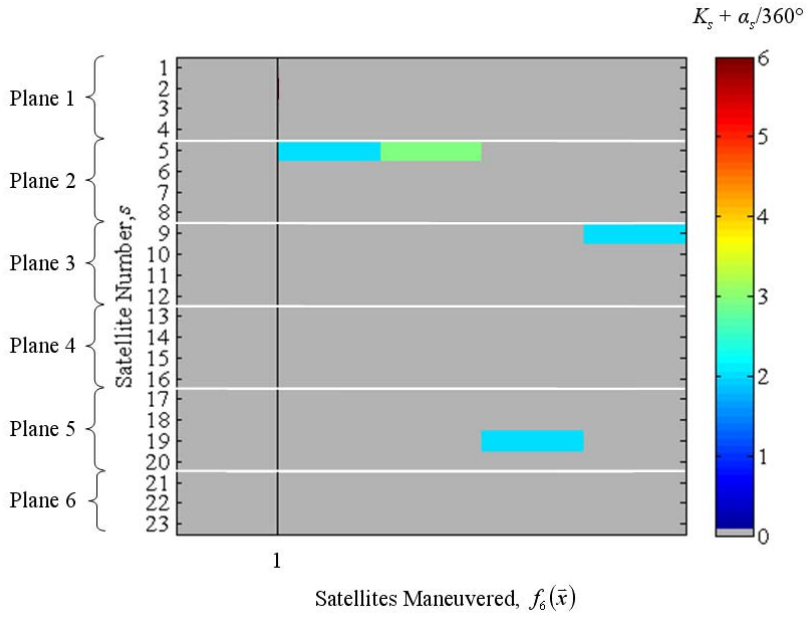


*Best coverage design from Case-1d-op in Section 5.3.2**,** Table 5-7

The decision variable plot for Case-1d-op-globe is compared to Case-1d-op in Fig. 5-47. **T**here are few similarities with regard to the satellites maneuvered for the global coverage case when compared to the hemispherical receiver grid assumption.



(a)



(b)

Figure 5-47: Hypervolume decision variable plot, Case-1d-op (a) global receiver grid, (b) original receiver grid, (one satellite lost, PDOP<2) independent phasing orbits, K_s, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\vec{x})$ ascending

There are two important points that should be taken away from this section. Increasing the grid to encompass more receivers will increase function evaluation time, but might make the problem easier to solve, ultimately reducing run time. One possible explanation for this observed fact was first discussed in Section 5.2. The global grid might have the effect of increasing the size of the feasibility islands, making it easier for the search algorithm to converge towards the fit regions of the search space in less time. The second, more obvious observation but nonetheless important, is to be very careful with the reconfiguration scenario assumptions regarding the behavior of access patterns. If an access constraint's impact on coverage is not well understood *a priori*, it is best to use the highest fidelity grid tolerable that will capture mission requirements. In the following concluding chapter, lessons like these will be summarized for this entire body of work and new research questions that have emerged as a result of this study will be posed.

Chapter 6

Conclusions, Reflections, and Potential Future Contributions

This work addressed the question of *how* an existing satellite constellation that has experienced the loss of capacity might be reconfigured to a new, *undetermined*, constellation using only the remaining vehicles. To answer this question, a *general* framework was developed that both approximates the set of Pareto-optimal reconfiguration options and guides a DM to the most preferred choice among multiple, competing objectives.

The objectives of the reconfiguration problem were classified as belonging to one of three categories: performance, cost, and risk (there could certainly be others and the framework allows for this). In this work, performance was associated with various metrics that assess a constellation's ability to 'cover' a receiver. Towards this end, the equations for the averaged rates of change of the classical elements, due to the first-order secular effects of J_2, were developed to model the position vector of a satellite at a future time. This vector was combined with minimum elevation and other constraints to produce an oscillatory access function that gave the rise and set intervals to a particular receiver. These intervals were then combined across multiple satellites to provide single- or multi-fold access statistics. From these intervals, coverage figures-of-merit were developed that provided, with a single quantity, an assessment of the performance of the *constellation* to the entire receiver grid. The specific metrics used for any given reconfiguration problem are mission dependent. In the case of GPS, the mission used to

demonstrate the framework, performance was assessed against two metrics, A-DVT and WCP-DVT, developed with guidance from the mission coverage specification.

The cost to maneuver is generally a tradeoff between the propellant required and time-of-flight to reach the future configuration. A variety of minimum energy solutions were introduced including those provided by Lambert, Gauss, and Battin as well as several less computationally intensive simplifications for coplanar and non-coplanar maneuvering. Again, the precise formulation for calculating energy or time-of-flight is mission dependent. For this work, the GPS constellation was constrained to approximately 12-hour precisely repeating ground traces precluding all but coplanar rephasing of the satellites. As such, the cost of maneuvering was quantified by the TTOF required to achieve the future constellation and the maximum energy or propellant expended by any of the maneuvered satellites. Another cost of reconfiguration identified in this work was the reduction in life expectancy incurred due to maneuvering and was quantified by calculating the propellant usage variance among the satellites maneuvered.

The final objective category, risk, can take many definitions. As it pertains to this work, however, the quantity of satellites that undergo a transfer determined the risk for reconfiguration. The minimum of this integer-valued function was sought, along with that of the TTOF, Worst ΔV or m_u, ΔV or m_u Var, and the maximum of the A-DVT and WCP-DVT. This six-objective tradeoff containing discrete, continuous, discontinuous, and nonlinear functions created a tradeoff landscape where deterministic optimization methods fail.

Building from recent contributions in satellite constellation design optimization and evolutionary computing research, a Master-Slave pMOEA, the ε-NSGA-2 was

selected and adapted to approximate the Pareto-optimal solutions in this complex fitness landscape. The ε-NSGA-2 uses epsilon-dominance archiving, auto-adaptive population sizing, and time continuation to collectively address the previously noted drawbacks of its parent algorithm, the NSGA-2. For this work, a shared parallel computing environment was used that consisted of a large combination of heterogeneous processing cores.

The original ε-NSGA-2 was designed to work efficiently on a dedicated small homogeneous cluster. In this research, several algorithm adaptations were made to improve the algorithm's scalability. The first modification relieved the master core of its duties to perform function evaluations. The second and more significant contributor towards mitigating the scalability issue was the development of asynchronous evolution wherein the master core did not wait for the final objective vector evaluations of a given population before proceeding to the next generation. Since the objective vector evaluated on the order of a minute per processing core and given that some cases required over 10 million function evaluations, a need emerged to ensure long duration run time. Whether the cause for premature termination was a node failure or reaching the maximum run time (95.0 hrs) in the shared resource environment, a check-pointing mechanism was developed that allowed for restarting of a given simulation where it left off.

Throughout this work, the framework has been referred to as *general* because this characteristic is what differentiates this research from others that came before it. Previously, work with the reconfiguration problem made specific and limiting assumptions about the future configuration in that it took the form of the well known solutions to the global coverage problem (Walker or SOC types). This assumption is

sufficient when more satellites are available for launch to increase capacity and restore global coverage or when the remaining vehicles have the ability to be maneuvered to the appropriate positions. These approaches, however, are inappropriate and potentially wasteful when global coverage is not required, and fail altogether when any possible reconfiguration is not capable of meeting a global coverage requirement. In this case, the optimization problem becomes one of discontinuous coverage, which is a much more difficult problem. The LC-ε-NSGA-2 was introduced into the framework as the optimization engine because it has the capacity to provide solutions to *both* types of coverage problems.

Another common drawback of previous work is the tendency to seek the optima of a single cost function through the subjective application of a preference structure prior to performing the optimization. In this work, however, the population-based approach to search and fitness based on a solution's epsilon-dominance and diversity allowed for the approximation of the entire set of optimal reconfiguration alternatives without an *a priori* introduction of preference information. The DMs preference structures were then applied, *a posteriori*, to the final non-dominated sets such that the cost of their choice was observed in the context other non-dominated solutions.

The framework was demonstrated on two loss scenarios for an idealized GPS constellation. The first considered the situation where a single satellite malfunctioned and the second asked how the constellation might be reconfigured after the loss of an entire plane. Glyph-hypervolume and decision-variable plots were produced in every case for both *idealized* and *operational* assumptions. The decision making process took a design-by-shopping approach where the most preferred designs in each hypervolume

were identified from a set of preference structures. In some cases, supplemental information including coverage contour plots or propellant-gage illustrations, were generated to aid with down-selection to a final choice.

Akin's Laws of Spacecraft Design[8] were developed by David Akin[9] to impart some of his wisdom gleaned during a lifetime of involvement in spacecraft and space systems design. Practitioners, professors, and students alike have found these bits of wisdom invaluable as they progressed through their careers. It is in a spirit similar to Akin that the following lessons-learned from the results of this work are distilled down into digestible pieces that might be enabling, or at the very least, prevent timely mistakes both in applications of the current framework or in future research:

- One should not assume that increasing the number of decision variables will produce a more challenging problem.
- The objectives of the reconfiguration problem pertaining to *coverage* have the greatest impact on the multimodality of the fitness landscape. As constraints like minimum elevation angle and PDOP are introduced, and more satellites are subtracted from the constellation, visibility is reduced and access gaps are created. As a result, there are more potential configurations to fill these gaps and hence longer run times.

---

[8] http://spacecraft.ssl.umd.edu/akins_laws.html, cited on September 17, 2008.

[9] Associate Professor of Aerospace Engineering, Department of Aerospace Engineering, University of Maryland

- Imposing operational assumptions can create infeasible regions in the fitness landscape, which can reduce the number of potential solutions that can populate the LC-ε-NSGA-2 archive, which can reduce run time.
- Avoid constraining the search with preconceived biases. Most of the results presented produced counter-intuitive designs. For example, the hypervolume for Case-2d-op did not contain a single design that maneuvered satellite 18, which had plenty of propellant available, while others with much less contributed to the reconfiguration.
- The hypervolumes can demonstrate that inaction after the loss of capacity is a better choice than any reconfiguration.
- When hundreds to thousands of non-dominated designs make up the hypervolume at run termination, a carefully selected set of preference structures applied *a posteriori* provides a rigorous methodology for reducing the set size.
- Multiple preference structures can lead to the identical non-dominated design, which demonstrates the preference induced dimensional reduction of one or more of the objectives.
- Choose the epsilon vector very carefully as they *do not* directly relate to tolerances. Suppose, for example, $k = 2$, and the epsilon vector is [1.0 1.0]. Depending on where the non-dominated designs fall in the grid, it is possible to have solutions nearly adjacent to each other in one dimension, well within the epsilon value of 1.0. This issue becomes exasperated as the number of dimensions increase.

- The use of an epsilon vector *does not* necessarily provide a free lunch with regard to reducing search time. When selecting an epsilon vector, a balance must be struck between an overly fine or coarse gird. The former can bring search to a crawl where little-to-no innovation takes place while the later will make search progress to a point and then *genetically drift* [115] among the large epsilon blocks.
- Clearly understand the impact of all access constraints (e.g., PDOP) on their associated access patterns for the entire region of interest before any assumptions are made that reduce the number of receivers to *simulate* the region.
- Do not assume any correlation between the size of the receiver grid and the difficulty of the problem. The global grid can actually make the problem easier to solve than one with a regional grid despite the increase in function evaluation time.

The limitations of the framework are detailed next to provide context for extensions and suggestions for future work. When the geometrical characteristics of the true Pareto-hypervolume are not known beforehand, solutions can be lost due to the chosen epsilon vector. In two-dimensions, these solutions would exist on segments of the Pareto-front that are almost horizontal or vertical. Methods [112] that dynamically adapt the epsilon vector have been proposed and demonstrated on test-functions; however, no work to-date has considered such schemes on real-world problems.

The GPS mission studied in this work had a repeat cycle of one day, which means that the access pattern for the constellation could be accurately captured by propagating

for 24 hours. Other constellations might have repeat cycles of a much greater duration requiring a much longer propagation and, in turn, an order of magnitude increase in function evaluation time. Additionally, the EOM developed for this research included the secular effects of J_2, but other missions might require modeling of third-body effects, drag, higher-order geopotential terms, etc. – all of which will result in extension to the function evaluation time. It is not difficult to imagine a problem that would require months of run time with the framework in its current form. For these types of problems, future work might consider how low fidelity models might be combined with high fidelity models to rapidly recognize inferior designs, reducing the total run time that would be required if only the high fidelity models were used.

One alternative solution to the high fidelity issue is the allocation of more parallel processing resources. This, however, is not a straightforward task. The simulations in this work used 250 cores from a cluster containing over 1200. Aging communication hardware and an interaction between the MPI implementation and the operating system resulted in frequent node failures and simulation aborts when more than 250 cores were tasked. Additional allocation of cores is also problematic on heavily shared resources because of the resulting simulation-execution wait times. With asynchronous evolution, there is no reason to believe that the efficiency of those processors tasked would decrease beyond 250 cores; however, other scalability bottlenecks will likely be encountered, such as the master not being able to keep up with slave coordination functions. Future work might focus on methods to mitigate these issues when greater than 250 cores are involved.

Despite these new challenges, the current framework is easily extensible to a variety of other closely related problems. In this work, constellations were reconfigured after the loss of a capacity. The inverse to this problem, or constellation build-up, could be studied by simply adapting the representation and applying the appropriate mission-specific objective functions. Since many satellite constellations are more than 10 years old, a more direct extension would allow for the reassessment of program requirements. If priorities or stakeholders have changed (and they almost always do over decade timescales), DMs could ask the question of how they might reconfigure the existing architecture to better meet new requirements. Regardless of the extension, the framework allows a DM to make informed and defensible decisions for an extremely difficult class of optimization problems that were otherwise computationally intractable.

References

1 Clark, A. C., "Extra-Terrestrial Relays Can Rocket Stations Give World-wide Radio Coverage," *Wireless World*, Oct. 1945, pp. 305–308.

2 Draim, J. E., "Satellite Constellations," *55th International Astronautical Congress,* Vancouver, Canada, 2004, IAC-04-A.5.01, pp 371–400.

3 Reichhart, T., "Satellite Smashers. Space Junk Will Soon Start Taking Out Expensive Spacecraft. What's the Cleanup Plan?," *Air and Space Magazine*, Feb./March 2008, pp. 50–53.

4 Butler, A., Bruno, M., Falghum, D., Doyle, J., "Ambiguous Intercept," *Aviation Week and Space and Technology*, Vol. 168, Iss. 8, Feb. 2008, p. 30.

5 Morring, F., Jr., "WORST EVER: China Asat Test Called Worst Single Debris Event Ever; Chinese anti-satellite test boosted space-debris population by 10% in an instant," *Aviation Week and Space Technology*, Vol. 166, Iss. 7, Feb. 2007, p. 20.

6 Robertson, B., and Stoneking, E., "Satellite GN&C Anomaly Trends," *Advances in the Astronautical Sciences*, Vol. 113, 2003, pp. 531–542.

7 Belov, A., Dorman, L., Iucci, N., Kryakunova, O., Ptitsyna, N., "The Relation of High- and Low-orbit Satellite Anomalies to Different Geophysical Parameters," *Effects of Space Weather on Technology Infrastructure*, Springer Netherlands, 2005, pp. 147–163.

8 Scialom, U., "Optimization of Satellite Constellation Reconfiguration," M.S. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, 2003.

9 Ahn, Y. T., and Spencer, D. B., "Optimal Reconfiguration of a Formation-Flying Satellite Constellation," *53rd International Astronautical Congress*, Houston, TX, Oct. 10-19, 2002, IAC-02-A.2.07.

10 Siddiqi, A., Mellein, J., de Weck, O., "Optimal Reconfigurations for Increasing Capacity of Communication Satellite Constellations," 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Austin, TX, April 18-21, 2005, AIAA-2005-2065.

11 Walker, J. G., "Circular Orbit Patterns Providing Continuous Whole Earth Coverage," Technical Report 70211, Royal Aircraft Establishment, United Kingdom, Nov. 1970.

12 Adams, W., and Rider, L., "Circular Polar Constellations Providing Continuous Single or Multiple Coverage Above a Specified Latitude," *The Journal of the Astronautical Sciences*, Vol. 35, No. 2, April-June 1987, pp. 155–192.

13 Ferringer, M., and Spencer, D., "Satellite Constellation Design Tradeoffs Using Multiple-Objective Evolutionary Computation," *Journal of Spacecraft and Rockets*, Vol. 43, 2006, pp. 1404-1411.

14 Ferringer, M., Clifton, R., Thompson, T., "Efficient and Accurate Evolutionary Multi-objective Optimization Paradigms for Constellation Design," *Journal of Spacecraft and Rockets*, Vol. 44, 2007, pp. 682–691.

15 Reed, P., Ferringer, M., Thompson, T., Kollat, J., "Parallel Evolutionary Multi-objective Optimization on Large Heterogeneous Clusters: An Applications Perspective," *AIAA Journal of Aerospace Computing Information and Communication*, accepted for publication May, 2008.

16 Deb, K., "Innovating Design Principles Through Optimization," *Proceedings of the 8th Annual Conference of Genetic and Evolutionary Computation*, Seattle WA, 2006, pp. 1629–1636.

17 Vallado, D., *Fundamentals of Astrodynamics and Applications,* 2nd ed., Microcosm Press, El Segundo, CA, 2001, pp. 1, 20–23, 304–350, 495, 575–582.

18 Beer, A., and Strand, A., *Kepler*, Pergamon Press, Oxford, 1975, p. 524.

19 Newton, I., *Mathematical Principles of Natural Philosophy*. Reprint edition, University of California Press, Berkley, CA, 1962 [1687], pp. 14, 415.

20 Wertz, J., *Mission Geometry: Orbit and Constellation Design and Management*, Microcosm Press, El Segundo, CA, Kluwer Academic Publishers, 2001, pp 61–65, 79–82, 469–497, 671–731.

21 Chao, C., *Applied Orbit Perturbation and Maintenance*, Aerospace Press, Los Angeles, CA, 2005.

22 Roy, A., *Orbital Motion*, 3rd ed., Institute of Physics Publishing, Bristol, 1988, pp. 312–319.

23 Prussing, J., and Conway, B., *Orbital Mechanics*, Oxford University Press, USA, 1993, pp. 26–29, 62–80.

24 Wiesel, W., *Spaceflight Dynamics*, McGraw-Hill, New York, NY, 1997, pp 57–60.

25 Chobotov, V., *Orbit Mechanics,* 3rd ed., AIAA Education Series, Reston, VA, Sept. 2002, pp. 7–9, 202–209.

26 Kozai, Y., "The Motion of a Close Earth Satellite," *The Astronomical Journal*, Vol. 64, No. 9, Nov. 1959.

27 Beer, A., and Strand, K., *Copernicus*, Oxford: Pergamon Press, 1975.

28 Lang, T., and Hanson, J., "Orbital Constellations which Minimize Revisit Time," AAS/AIAA Astrodynamics Specialist Conference, Lake Placid, NY, AAS Paper 83-402, Aug. 22-25, 1983.

29 Williams, E., Crossley, W., Lang, T., "Average and Maximum Revisit Time Trade Studies for Satellite Constellations using a Multi-Objective Genetic Algorithm," *Advances in the Astronautical Sciences*, Vol. 105, Jan. 2000, p. 653.

30 Draim, J., "Three- and Four-satellite Continuous Coverage Constellations," AIAA/AAS Astrodynamics Specialist Conference, Seattle, WA, AIAA-84-1996, Aug. 20-22, 1984.

31 Lüders, R., "Satellite Networks for Continuous Zonal Coverage," *American Rocket Society Journal*, Vol. 31, No. 2, 1961, pp. 179–184.

32 Smith, J., "Application of Optimization Techniques to the Design and Maintenance of Satellite Constellations," M.S. Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, June, 1999.

33 George, E., "Optimization of Satellite Constellations for Discontinuous Global Coverage via Genetic Algorithms," *Advances in the Astronautical Sciences*, Vol. 97, Aug. 1997, pp. 333–346.

34 Ely, T., Crossley, W., Williams, E., "Satellite Constellation Design for Zonal Coverage Using Genetic Algorithms," *Advances in the Astronautical Sciences*, Vol. 99, Feb. 1998, pp. 443–460.

35 Lang, T., "A Parametric Examination of Satellite Constellations to Minimize Revisit Time for Low Earth Orbits Using a Genetic Algorithm," *Advances in the Astronautical Sciences*, Vol. 109, Aug. 2001, pp. 625–640.

36 Mason, W., Coverstone-Carroll, V., Hartmann, J., "Optimal Earth Orbiting Satellite Constellations via a Pareto Genetic Algorithm," AIAA Paper 98-4381, Aug. 1998.

37 Crossley, W., and Williams, E., "Simulated Annealing and Genetic Algorithm Approaches for Discontinuous Coverage Satellite Constellations Design," *Engineering Optimization*, Vol. 32, No. 3, 2000, pp. 353–371.

38 Escobal, P., *Methods of Orbit Determination*, John Wiley and Sons, New York, 1965, pp. 188–197.

39 Bate, R, Mueller, D., White, J., *Fundamentals of Astrodynamics*, Dover Publications, New York, 1971, pp. 193–198.

40 Battin, R., *An Introduction to the Mathematics and Methods of Astrodynamics*, AIAA Education Series, New York, 1987, pp 325–342.

41 Ferringer, M., Spencer, D., Reed, P., Clifton, R., Thompson, T., "Pareto-hypervolumes for the Reconfiguration of Satellite Constellations," *AIAA Journal*, submitted for publication June, 2008.

42 Thompson, R., "Calculating Response Time Statistics", Aerospace Technical Memorandum, N97-5040-RCT-003, July 18, 1997, pp 1–6.

43 Goldberg, D., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Publishing Inc., Reading, Massachusetts, 1989.

44 Fletcher, R., *Practical Methods of Optimization, Vol. 1, Unconstrained Optimization*, John Wiley and Sons Inc., New York, 1981.

45 Escobal, R., *Methods of Astrodynamics*, John Wiley and Sons, New York, 1968, p. 66.

46 Bryson, A., and Ho, Y., *Applied Optimal Control*, Hemisphere Publishing Corporation, Washington D.C., 1987.

47 Pareto, V., *Manuale di Economia Politica*, Societa Editrice Libraria, Milano, Italy, 1906, translated to English by A. S. Schwier as, *Manual of Political Economy*, MacMillan, New York, 1971.

48 Coello Coello, C., Van Veldhuizen, D., Lamont, G., *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer Academic Publishers, New York, NY, 2002, pp. 14–15.

49 Deb, K., *Multi-Objective Optimization using Evolutionary Algorithms*, New York, NY, John Wiley and Sons LTD, 2001.

50 Cantor, G., "Contributions to the Foundation of Transfinite Set Theory," *Mathematische Annalen*, Vol. 46, 1895, pp. 481–512.

51 Cohon, J., and Marks, D., "A Review and Evaluation of Multi-Objective Programming Techniques," *Water Resources Research*, Vol. 11, No. 2, 1975, pp. 208–220.

52 Zadeh, L., "Optimality and Nonscalar-Valued Performance Criteria," *IEEE Transactions on Automatic Control*, Vol. 8, No. 1, 1963, pp. 59–60.

53 Stadler, W., "Preference Optimality and Applications to Pareto Optimality," In Leitmann, G., Marzello, A,. eds., *Multi-criteria Decision Making*, Vol. 211, Springer-Verlag, New York, NY, 1975.

54 Osyczka, A., *Multicriteria Optimization in Engineering with FORTRAN Programs*, Ellis Horwood Limited, West Sussex, UK, 1984.

55 Charnes, A., and Cooper, W., *Management Models and Industrial Applications of Linear Programming*, Vol. 1, John Wiley, New York, NY, 1961.

56 Ijiri, Y., *Management Goals and Accounting for Control*, North-Holland, Amsterdam, 1965.

57 Gembicki, F., *Vector Optimization for Control with Performance and Parameter Sensitivity Indices*, PhD Thesis, Department of Systems Engineering, Case Western Reserve University, Cleveland, Ohio, 1974.

58 Rao, S., "Multiobjective Optimization in Structural Design with Uncertain Parameters and Stochastic Processes," *AIAA Journal*, Vol. 22, No. 11, 1984, pp. 1670–1678.

59 von Neumann, J., and Morgenstern, O., *Theory of Game and Economic Behavior*, Princeton University Press, Princeton, New Jersey, 2nd ed., 1947.

60 Gass, S., and Saaty, T., "The Computational Algorithm for the Parametric Objective Function," *Naval Research Logistics Quarterly*, Vol. 2, 1955, pp. 39–45.

61 Hwang, C., Paidy, S,. Yoon, K., "Mathematical Programming with Multiple Objectives: A Tutorial," *Computing and Operational Research*, Vol. 7, 1980, pp. 5–31.

62 Goicoechea, A., Duckstein, L., Fogel, M., "Multi-objective Programming in Watershed Management: A Study of the Charleston Watershed," *Water Resources Research*, Vol. 12, No. 6, 1976, pp. 1085–1092.

63 Cohon, J., *Multiobjective Programming and Planning*, Academic Press, Burlington, MA, 1978.

64 Monarchi, D., Kisiel, C., Duckstein, L., "Interactive Multiobjective Programming in Water Resources: A Case Study," *Water Resources Research*, Vol. 9, No. 4, 1973, pp 837–850.

65 Eberhart, R., and Kennedy, J., "A New Optimizer using Particle Swarm Theory," *Proceedings of the Sixth International Symposium on Micromachine and Human Science*, Nagoya, Japan, 1995, pp. 39–43.

66 Dorigo, M., Stützle, T., *Ant Colony Optimization*, MIT Press, Cambridge, MA, 2004.

67 Schaffer, J., *Some Experiments in Machine Learning Using Vector Evaluated Genetic Algorithms*, Ph.D. Thesis, Department of Computer Science, Vanderbilt University, Nashville, TN, Dec. 1984.

68 Robins, H., "Some Aspects of the Sequential Design of Experiments," In *Bulletin of the American Mathematical Society*, Vol. 55, 1952, pp. 527–535.

69 Holland, J., *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor, Michigan, 1975.

70 Schwefel, H., *Evolution and Optimum Seeking*, John Wiley and Sons Inc., New York, NY, 1995.

71 Dowsland, K., "Simulated Annealing," In Reeves, C., ed., *Modern Heuristic Techniques for Combinatorial Problems*, Ch. 2, John Wiley and Sons Inc., New York, NY, 1993, pp 20–69.

72 Fonseca, C., and Flemming, P., "Genetic Algorithms for Multiobjective Optimization: Formulation, Discussion, and Generalization," *Proceedings of the Fourth International Conference on Evolutionary Programming*, 1993, pp. 416–423.

73 Srinivas, N., and Deb, K., "Multi-objective Function Optimization using Non-dominated Sorting Genetic Algorithms", *Evolutionary Computation*, Vol. 2, No. 3, 1994, pp. 221–248.

74 Horn, J., Nafploitis, N., Goldberg, D., "A Niched Pareto Genetic Algorithm for Multi-objective Optimization," *Proceedings of the First IEEE Conference on Evolutionary Computation*, 1994, pp. 82–87.

75 Zitzler, E., and Thiele, L., "An Evolutionary Algorithm for Multiobjective Optimization: The Strength Pareto Approach," Technical Report 43, Computer Engineering and Networks Laboratory, Swiss Federal Institute of Technology, Zürich, Switzerland, 1998.

76 Deb, K., Pratap, A., Agarwal, S., Myarivan, T., "A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, Vol. 6, April 2002, pp. 182–197.

77 Deb, K., and Agrawal, R., "Simulated Binary Crossover for Continuous Search Space," *Complex Systems*, Vol. 9, No. 2, 1995, pp. 115–148.

78 Kollat, J., and Reed, P., "The Value of Online Adaptive Search: A comparison of NSGA-II, ε-NSGAII, and εMOEA," *Evolutionary Multi Criterion Optimization: Third International Conference (EMO 2005)*, Guanajuato, Mexico, 2005, pp. 386–398.

79 Tang, Y., Reed, P., Wagener, T., "How Efficient and Effective are Evolutionary Multi-objective Algorithms at Hydrologic Model Calibration?," *Hydrology and Earth System Sciences,* Vol. 10, 2006, pp. 289–307.

80 Kollat, J., and Reed, P., "Comparing State-of-the-Art Evolutionary Multi-Objective Algorithms for Long-Term Groundwater Monitoring Design," *Advances in Water Resources,* Vol. 29, 2006, pp. 792–807.

81 Laumanns, M., Thiele, L., Deb, K., Zitzler, E., "Combining Convergence and Diversity in Evolutionary Multiobjective Optimization," *Evolutionary Computation*, Vol. 10, 2002, pp. 263–282.

82 Goldberg, D., *The Design of Innovation: Lessons from and for Competent Genetic Algorithms*, Kluwer Academic Publishers, Norwell, MA, 2002, pp. 220–228.

83 Tang, Y., Reed, P., Kollat, J., "Parallelization Strategies for Rapid and Robust Evolutionary Multi-objective Optimization in Water Resources Applications," *Advances in Water Resources*, Vol. 30, No. 3, 2007, pp. 335–353.

84 Spall, J., *Introduction to Stochastic Search and Optimization*, John Wiley and Sons, Hoboken, New Jersey, 2003.

85 Van Veldhuizen, D., Zydallis, J., Lamont, G., "Considerations in Engineering Parallel Multiobjective Evolutionary Algorithms," *IEEE Transactions on Evolutionary Computation*, Vol. 7, No. 2, April 2003, pp. 144–173.

86 Cantu-Paz, E., *Efficient and Accurate Parallel Genetic Algorithms,* Kluwer Academic Publishers, Boston, 2000, Chapters 4-7.

87 Darwin, C., *Journal of Researches into the Natural History and Geology of the Countries Visited During the Voyage Round the World of H.M.S. Beagle*, 11th ed., John Murray, London, 1913, p. 404.

88 Mayr, E., "Speciational Evolution or Punctuated Equilibria," In Somit, A, Peterson, S., *The Dynamics of Evolution*, Cornel University Press, New York, 1992, pp 25–26.

89 Dawkins, R., *The Blind Watchmaker*, W. W. Norton, New York, NY, 1996, p. 227.

90 Gropp, W., *Using MPI: Portable Parallel Programming with Message-Passing Interface*, 2nd ed., The MIT Press, Cambridge, MA, 1999.

91 Snir, M., Otto, S., Huss-Lederman, S., Walker, D., Dongarra, J., *MPI: The Complete Reference*, The MIT Press, Cambridge, MA, 1966.

92 Damani, O., and Garg, V., "Fault-Tolerant Distributed Simulation," *Proceedings of the 12th Workshop on Parallel and Distributed Simulation*, Banff, Alberta, Canada, 1998, pp. 38–45.

93 Yinger, C., "Operation and Application of the Global Positioning System," *Crosslink*, Summer 2002, The Aerospace Corporation, El Segundo, CA, pp. 12–16.

94 Kaplan, E., ed., *Understanding GPS: Principles and Applications*, Artech House Publishers, Norwood, MA, Feb. 1996, pp. 59–61.

95 Misra, P., and Enge, P., *Global Positioning System Signals, Measurements, and Performance*, Ganga-Jamuna Press, Lincoln, Massachusetts, 2001.

96 United States Department of Homeland Security, United States Coast Guard, "GPS Positioning System Standard Positioning Service Signal Specification, Annex A, Standard Positioning Service Performance Specification," 2nd ed., June 2, 1995, p. A-1.

97 Massatt P., and Brady, W., "Optimizing Performance Through Constellation Management," *Crosslink*, Summer 2002, The Aerospace Corporation, El Segundo, CA, pp. 17–21.

98 Berger, J., *Statistical Decision Theory and Bayesian Analysis*, 2nd ed., Springer-Verlag, New York, NY, 1980.

99 Sutton, R., and Barto, A., *Reinforcement Learning an Introduction*, The MIT Press, Cambridge, MA, 1998, pp. 4–5, 130, 145, 236–238.

100 Wolpert, H., and Macready, W., "No Free Lunch Theorems for Optimization," *IEEE Transactions on Evolutionary Computation*, Vol. 1, No. 1, April, 1997, pp 67–82.

101 Darwin, C., *On the Origin of Species by Means of Natural Selection or the Preservation of Favoured Races in the Struggle for Life*, Murray, London, UK, 1859, Chapter 6.

102 Wertz, J., and Larson, W., *Space Mission Analysis and Design*, Springer, New York, NY, 1999, p. 688.

103 Hatch, R., Jung, J., Enge, P., Pervan, B., "Civilian GPS: The Benefits of Three Frequencies," *GPS Solutions*, Vol. 3, No. 4, 2000.

104 Lachapelle, G., Cannon, M., Bird, J., "Evaluation of GPS-Aided Artillery Position and Orientation Methods," *Journal of Military and Strategic Studies*, Vol. 1, No. 1, 1998.

105 Davis, B., AuYeung, M., Clark, J., Lee, C., Palko, J., and Thomas, M., "Reflections on Building a High-performance Computing Cluster using FreeBSD," in *BSDCan*, University of Ottawa, 2007.

106 Sanjuan, R., and Nebot, M., "A Network Model for the Correlation between Epistasis and Genomic Complexity," *PLoS ONE Journal*, Vol. 3, No. 7, 2008.

107 Byron, M., ed., *Satisficing and Maximizing Moral Theorists on Practical Reason*, Cambridge University Press, Cambridge, UK, 2004.

108 Balling, R,. "Design by Shopping: A New Paradigm," *Proceedings of the Third World Congress of Structural and Multidisciplinary Optimization*, Buffalo, NY, 1999, pp. 295–297.

109 Stump, G., Yukish, M., Simpson, T., O'Hara, J., "Trade Space Exploration of Satellite Datasets Using a Design by Shopping Paradigm," *Proceedings of the IEEE Aerospace Conference*, Vol. 6, March 6-13, 2004, pp. 3885–3895.

110 Easton, R., and Brescia, R., "Continuously Visible Satellite Constellations," Report on NRL Progress, July 1969, pp. 1–5.

111 Kreyszig, E., *Advanced Engineering Mathematics*, 8th ed., John Wiley and Sons Inc., New York, NY, p. 841.

112 Hernandez-Diaz, A., Santana-Quintero, L., Coello Coello, C., Molina, J., "Pareto-adaptive ε-dominance," *Evolutionary Computation*, Vol. 15, No. 4, pp 493–517.

113 Deb, K., Thiele, L., Laumanns, E., Zitzler, E., "Scalable Test Problems for Evolutionary Multi-objective Optimization," *Proceedings of the Congress on Evolutionary Computation*, Vol. 1, May 2002, pp. 825-830.

114 Kollat, J., and Reed, P., "A Computational Scaling Analysis of Multiobjective Evolutionary Algorithms in Long-Term Groundwater Monitoring Applications," *Advances in Water Resources*, Vol. 30, No. 3, pp 408-419.

115 Thierens, D., Goldberg, D., Pereira, A., "Domino Convergence, Drift, and the Temporal-salience Structure of Problems," *Evolutionary Computation Proceedings*, IEEE World Congress on Computational Intelligence, May, 1998, pp. 535-540.

Appendix A

Nomenclature

A.1 Abbreviations

A	offline archive size containing epsilon non-dominated solutions
a	semimajor axis, km
a_e	semimajor axis of the earth, km
a_{phase}	semimajor axis of phasing orbit, km
a_{rep}	semimajor axis of repeating ground track orbit, km
$\overline{a}$	mean semimajor axis, km
$\overline{a}_0$	mean semimajor axis at epoch, km
B	inequality constraints
C	folds array
c	connected runs
D	decision variables
d	integral number of days
d_{crowd}	crowding distance
d_b	length of cuboid base
d_h	length of cuboid height
E	eccentric anomaly, deg or rad
e	eccentricity

$\bar{e}$	mean eccentricity
$\bar{e}_0$	mean eccentricity at epoch
F	folds of coverage
f_{oscil}	oscillatory access function
G	gravitational constant, $\text{km}^3/(\text{kg}\cdot\text{sec}^2)$
g_0	gravitational constant at sea level, km/sec^2
H	total accesses
h	access interval number
I_{sp}	specific impulse, sec
i	inclination, deg
$\bar{i}$	mean inclination, deg
$\bar{i}_0$	mean inclination at epoch, deg
J_2	oblateness term in geopotential expansion
j	inequality constraint index
K	integer phasing revolutions
K_s	integer phasing revolutions for satellite, s
k	dimension of the objective vector
L	total receivers in the target deck
l	receiver index
M	mean anomaly, rad
M_0	mean anomaly at epoch, rad
$\overline{M}$	mean, mean anomaly, rad

$\overline{M}_0$ mean, mean anomaly at epoch, rad

m mass, kg

m_a propellant mass available before a burn, kg

m_d propellant mass depleted before the maneuver, kg

m_e mass of the earth, kg

m_f satellite mass after a burn, kg

m_i satellite mass before a burn, kg

m_r propellant mass remaining after a burn, kg

m_{sat} mass of the satellite, kg

m_u propellant mass used during a burn, kg

N population size

N_v variable length population size

n mean motion, rad/sec

n_{rep} estimated mean motion for repeat ground track, rad/sec

$\overline{n}$ perturbed mean motion, rad/sec

P orbital period, sec

P_2 Legendre polynomial of order 2

P^* Pareto-optimal set

p integral number of orbit periods

q repeat cycle ratio

R disturbing potential, km^2/sec^2

$\widetilde{R}$ secular disturbing potential, km^2/sec^2

R_{avg} average disturbing potential over one revolution, km^2/sec^2

r magnitude of satellite position vector, km

r_e magnitude of the earth's radius, km

S total satellites transferred

s satellite index

SOC_T number of satellite in a Streets-Of-Coverage plane

SOC_P number of planes in a Streets-Of-Coverage constellation

T inequality constraint

t time from epoch, sec

t_0 epoch time, sec

t_{rise_h} access interval rise time for access h, sec from epoch

t_{set_h} access interval set time for access h, sec from epoch

t_{sid} time in a sidereal day, sec

t_{tran} maneuver transfer time, sec

U aspherical gravitational potential, km^2/sec^2

U_{sphere} spherical gravitational potential, km^2/sec^2

u equality constraint index

V equality constraint

W_T total number of satellites in a Walker Delta Pattern

W_P total number of planes in a Walker Delta Pattern

W_F relative spacing between satellites in adjacent planes in a Walker Delta Pattern

w equality constraints

x,y,z satellite position variables in the $\hat{I}\hat{J}\hat{K}$ reference frame

A.2 Symbols

α phase angle, deg

γ vernal equinox

ΔV change in velocity, km/sec

ε epsilon level for epsilon-constraint method

ε_k epsilon level for objective function k

Λ objective space

λ longitude, deg

μ gravitational parameter, km^3/sec^2

ν true anomaly, deg

$\dot{\nu}$ true anomaly rate, deg/sec

ξ satellite elevation angle, deg

$\xi_{\min}$ satellite minimum elevation angle, deg

ρ Position Dilution Of Precision constraint

ϕ argument of latitude, deg

ϕ_{gc} geocentric latitude, deg

Ψ design space

Ω right ascension of the ascending node, deg

$\overline{\Omega}$ mean right ascension of the ascending node, deg

$\overline{\Omega}_0$ mean right ascension of the ascending node at epoch, deg

ω argument of perigee, deg

$\overline{\omega}$ mean argument of perigee, deg

$\overline{\omega}_0$ mean argument of perigee at epoch, deg

ω_e Earth's rotation rate, deg/sidereal sec

ζ substitution variable

A.3 Vectors

$\vec{f}$ objective function vector

$\vec{F}$ force, N

$\vec{F}_g$ force of gravity, N

G geometry matrix

$\hat{I}\hat{J}\hat{K}$ Earth Centered Inertial coordinate system

$\vec{O}$ observation vector from the satellite to the receiver, km

$\hat{O}$ unit observation vector from the satellite to the receiver

$\hat{P}\hat{Q}\hat{W}$ perifocal coordinate system

$\vec{R}$ receiver vector from the center of the earth to a receiver, km

$\hat{R}$ unit receiver vector from the center of the earth to a receiver

ROT elementary rotation matrix

$\vec{r}$ satellite position vector relative to $\hat{I}\hat{J}\hat{K}$ coordinate system, km

$\ddot{\vec{r}}$ acceleration, km^3/sec^2

$\vec{r}_e$ Earth position vector in the $\hat{X}\hat{Y}\hat{Z}$ coordinate system, km

$\vec{r}_{ECEF}$ satellite position vector in the earth centered earth fixed coordinate system, km

$\ddot{\vec{r}}_e$ Earth acceleration vector in the $\hat{X}\hat{Y}\hat{Z}$ coordinate system, km/sec^2

$\vec{r}_{\text{sat}}$ satellite position vector in the $\hat{X}\hat{Y}\hat{Z}$ coordinate system, km

$\ddot{\vec{r}}_{\text{sat}}$ satellite acceleration vector in the $\hat{X}\hat{Y}\hat{Z}$ coordinate system, km/sec^2

$\vec{V}$ most preferred design preference vector

$\vec{W}$ preference articulation vector

$\hat{X}\hat{Y}\hat{Z}$ inertial coordinate system

$\vec{x}$ decision variable vector

$\vec{x}^*$ decision variable vector that is a member of P^*

Appendix B

Acronyms

A-DVT	Average Daily Visibility Time
DMs	Decision Makers
DVT	Daily Visibility Time
ε-NDS	Epsilon Non-Dominated Solutions
ε-NSGA-2	Epsilon Non-Dominated Sorting Genetic Algorithm Two
EA	Evolutionary Algorithm
EC	Evolutionary Computing
ECEF	Earth-Centered Earth-Fixed
ECI	Earth Centered Inertial
EOM	Equations Of Motion
FDVT	Multi-fold Daily Visibility Time
GA	Genetic Algorithms
GEO	Geosynchronous Earth Orbit
GPS	Global Positioning System
LC-ε-NSGA-2	Large-Cluster Epsilon Non-Dominated Sorting Genetic Algorithm Two
LCW	Linear Combination of Weights
LEO	Low Earth Orbit
MOEA	Multi-Objective Evolutionary Algorithm
MOGA	Multi-Objective Genetic Algorithm

MOO	Multi-Objective Optimization
MOPs	Multi-Objective Optimization Problems
MP	Multiple-Population
MS	Master-Slave
NPGA	Niched-Pareto Genetic Algorithm
NSGA	Non-dominated Sorting Genetic Algorithm
NSGA-2	Non-dominated Sorting Genetic Algorithm Two
PDOP	Position Dilution Of Precision
pMOEA	parallel Multi-Objective Evolutionary Algorithm
SDT	Statistical Decision Theory
SOC	Streets-Of-Coverage
SGE	Sun Grid Engine
SPEA	Strength-Pareto Evolutionary Algorithm
TTOF	Total-Time-Of-Flight
US	United States
VEGA	Vector Evaluated Genetic Algorithm
VTS	Vandenberg Telemetry Station
WCP-DVT	Worst-Case-Point Daily Visibility Time

Appendix C

Calculating $\vec{r}$ in a Inertial Reference Frame from Classical Orbital Elements

In Chapter 2, Kozai's analytical method is used to provide averaged orbital elements as a function of time for the purposes of determining the future state of a satellite. The oscillatory function used to calculate accesses requires the satellite's position vector, $\vec{r}$, in the ECI reference frame. This appendix provides the algorithm for converting the averaged elements $\left(\overline{a},\overline{e},\overline{i},\overline{\Omega},\overline{\omega},\overline{M}\right)$ into $\vec{r}$. In this Appendix, the over-bar notation for the classical elements is dropped for readability, but averaged elements are implied.

Note, the following algorithm [17] for converting classical elements to a position vector in the ECI frame only holds true for the circular orbit assumed for the mission discussed in this research. Noting that ω = 0.0 for circular orbits, the position vector in the perifocal ($\hat{P}\hat{Q}\hat{W}$) coordinate system,

$$\vec{r}_{\hat{P}\hat{Q}\hat{W}} = \begin{bmatrix} a(\cos(M)) \\ a(\sin(M)) \\ 0 \end{bmatrix}, \tag{C.1}$$

is calculated. Next, this position vector must be converted from the perifocal frame to the ECI frame. This is accomplished using two elementary rotation matrices, *ROT1, ROT2*.

$$ROT1(\xi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\zeta) & -\sin(\zeta) \\ 0 & \sin(\zeta) & \cos(\zeta) \end{bmatrix} \tag{C.2}$$

$$ROT2(\xi) = \begin{bmatrix} \cos(\zeta) & -\sin(\zeta) & 0 \\ \sin(\zeta) & \cos(\zeta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{C.3}$$

The position vector in the ECI frame,

$$\vec{r}_{\hat{I}\hat{J}\hat{K}} = [ROT3(\Omega)][ROT1(i)]\vec{r}_{\hat{P}\hat{Q}\hat{W}}, \tag{C.4}$$

is found by pre-multiplying the position vector in the perifocal frame by the first rotation matrix then by the second, replacing the substitution variable, ζ, by the inclination, i, and right ascension, Ω respectively. Note that any appearance of $\vec{r}$ in the body of this work is the position vector in the ECI frame, $\vec{r}_{\hat{I}\hat{J}\hat{K}}$.

Appendix D

Calculating $\vec{R}$ in a Inertial Reference Frame from Receiver Latitude and Longitude

Each point in the receiver grid is specified by geocentric latitude, ϕ_{gc}, and longitude, λ, pairs. With the goal of calculating visibility to these locations, the receiver vector, $\vec{R}$, must be specified in the ECI frame for incorporation into the oscillatory access function. Latitude and longitude, however, are a type of Earth-Centered Earth-Fixed (ECEF) representation (a coordinate system that rotates with the Earth). The first step in the transformation is to form an ECEF position vector, $\vec{R}_{ECEF}$,

$$\vec{R}_{ECEF} = \begin{bmatrix} r\cos(\phi_{gc})\cos(\lambda) \\ r\cos(\phi_{gc})\sin(\lambda) \\ r\sin(\phi_{gc}) \end{bmatrix}, \tag{D.1}$$

from ϕ_{gc} and λ. Next, the ECEF vector of Eq. D.1 is converted to the inertial vector,

$$\vec{R}_{\hat{I}\hat{J}\hat{K}} = \begin{bmatrix} \cos(\omega_e t) & -\sin(\omega_e t) & 0 \\ \sin(\omega_e t) & \cos(\omega_e t) & 0 \\ 0 & 0 & 1 \end{bmatrix} \vec{R}_{ECEF}, \tag{D.2}$$

in ECI at any time, t, by premultiplying $\vec{R}_{ECEF}$ by a rotation matrix about the polar $\hat{K}$ axis (see Eq. C.3 in Appendix C). The argument of the rotation matrix is the product of t and the Earth's rotation rate, ω_e.

Appendix E

Root Finding Methods

Recall from Section 2.2.1 that the roots of an oscillatory access function provide the rise and set times of a satellite to a particular receiver location. The two methods used in this research to determine roots, Newton's or bisection, are described in the following sections. The continuous polynomial,

$$f(t) = t^4 - t^3 - 7t^2 + t + 6 \qquad t \in [2,4], \tag{E.1}$$

simulates an oscillatory access function (illustrated in Fig. E-1) and is used to demonstrate root finding with each method. The goal is to find the time, t, during the time step bounded by $[a,b]$ where the function produces a root.

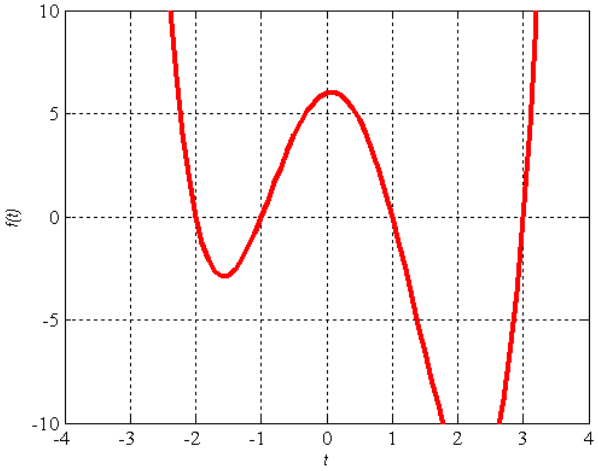


Figure E-1: Oscillatory function example

E.1 Newton's Method

Newton's method can be derived from a Taylor series expansion of $f(t)$ about a point $t = t_0 + \chi$, where χ is the desired accuracy of the sought after root [111]. Using only the first order terms in the series expansion,

$$t_{n+1} = t_n - \frac{f(t_n)}{f'(t_n)}, \; n = 0, 1, 2, \ldots, \tag{E.2}$$

iteration proceeds from the initial guess, t_0, until the error is less than χ.

In order to use this method, at the very least, the first derivative of the oscillatory function is required, the function must be continuous on $[a,b]$, and a good starting point for the initial guess is generally helpful. Table E-1 illustrates the iteration for Eq. E.1 with $t_0 = 2.5$ and $\chi = .001$. Newton's method finds the root, 3.0, after 5 iterations.

Table E-1: Newton's Method iteration

Iteration	t_n	$f(t_n)$	$f'(t_n)$
0	2.5	−11.8125	9.75
1	3.711538	51.91944	112.2254
2	3.248903	12.48378	61.02274
3	3.044327	1.848711	43.43405
4	3.001763	0.070656	40.13412
5	3.000003	0.000118	40.00022

E.2 Bisection

When Newton's method diverges or derivative information is not available, the bisection method is used. For this method to be successful, continuity is still a requirement and $f(a)$ and $f(b)$ must be of opposite signs. The bisection method divides the interval $[a,b]$ in half to produce c. There are two possibilities: the first is that $f(a)$ and $f(c)$ have opposite signs; the second is that $f(b)$ and $f(c)$ have opposite signs. Bisection is applied recursively to the subinterval, either $[a,c]$ or $[b,c]$ where the sign change occurs, until a particular convergence tolerance is achieved. Table E-2 illustrates iteration with bisection on Eq. E.1 where $a = 2.5$ and $b = 4$. The method converges in 13 iterations.

Table E-2: Bisection Method iteration

Iteration	a	b	c	$f(a)$	$f(b)$	$f(c)$
0	2.5	4	3.25	−11.8125	90	12.5508
1	2.5	3.25	2.875	−11.8125	12.5508	−4.4275
2	3.2	2.875	3.0625	12.5508	−4.4275	2.6511
3	2.875	3.0625	2.9688	−4.4275	2.6511	−1.2132
4	3.0625	2.9688	3.0156	2.6511	−1.2132	0.6343
5	2.9688	3.0156	2.9922	−1.2132	0.6343	−0.3102
6	3.0156	2.9922	3.0039	0.6343	−0.3102	0.1568
7	2.9922	3.0039	2.998	−0.3102	0.1568	−0.078
8	3.0039	2.998	3.001	0.1568	−0.078	0.0391
9	2.998	3.001	2.9995	−0.078	0.0391	−0.0195
10	3.001	2.9995	3.0002	0.0391	−0.0195	0.0098
11	2.9995	3.0002	2.9999	−0.0195	0.0098	−0.0049
12	3.0002	2.9999	3.0001	0.0098	−0.0049	0.0024
13	2.9999	3.0001	3	−0.0049	0.0024	−0.0012

The bisection method converges linearly, which is quite slow relative to Newton's (which is at least quadratic), but is used when needed because convergence is guaranteed. It is assumed that the time step selected is small enough (one-tenth of the orbital period) such that multiple roots will not exist in [a,b].

Appendix F

Multi-fold Access Interval Algorithm

The GPS constellation requires a minimum of four-fold coverage to any receiver location in order to perform its mission. The algorithm used to calculate multi-fold access is detailed in this appendix.

Recall from Section 2.2.1 that access interval arrays consist of t_{rise} and t_{set} pairs that mark the times when the satellite rises into view of the receiver and sets out of view during the propagation. In order to calculate multi-fold coverage, a *folds array*, *C*, is constructed. *C* is a multi-dimensional array of access intervals (which are themselves arrays) with the number of columns equal to the number of satellites in the constellation and number of rows equal to the maximum required fold calculation. An example of the structure of this array for the GPS constellation with 24 satellites requiring four-fold coverage is shown in Table F-1. The individual array elements in Table F-1 are initialized to zero interval arrays ([t_{rise} = 0 t_{set} = 0]).

Table F-1: Initialization of the folds array, *C*

	Satellite Identifier					
Folds	1	2	3	4	…	24
1	[0 0]	[0 0]	[0 0]	[0 0]	[0 0]	[0 0]
2	[0 0]	[0 0]	[0 0]	[0 0]	[0 0]	[0 0]
3	[0 0]	[0 0]	[0 0]	[0 0]	[0 0]	[0 0]
4	[0 0]	[0 0]	[0 0]	[0 0]	[0 0]	[0 0]

In order to populate *C*, the individual access intervals for every satellite must be calculated first. Section 2.2.1 describes how this is accomplished via the oscillatory access function. Hereafter, the *sets* of access intervals for each satellite to a single ground point will be delineated by the satellite identifier enclosed in square brackets, [s_1]. Table F-2, which is used to describe the multi-fold access interval algorithm, is populated from left to right, with the rightmost column containing the one-fold to four-fold access arrays for the constellation. The algorithm begins by calculating element (1,1), denoted by [1Fs_1] which is simply equal to the access interval array for satellite 1, [s_1]. Next, moving to element (1,2), the single-fold access array for satellites 1 and 2, [1Fs_2], is calculated by taking the *union* of [s_1] and [1Fs_1]. In element (2,2), the two-fold access array for satellites 1 and 2, [2Fs_2], is found by *intersecting* the satellite 2 access array, [s_2], with [1Fs_1] calculated previously. Element (1,3), the single-fold access for satellites 1, 2, and 3, [1Fs_3] is calculated by taking the *union* of the satellite 3 access array, [s_3] and the previously calculated [1Fs_2]. Moving down column 3 to element (2,3), the two-fold access interval array for satellites 1, 2, and 3, [2Fs_3], is calculated by first *intersecting* the access array for one-fold coverage by satellites 1 and 2, [1Fs_2], with the access array of satellite 3, [s_3]. This result is then *unioned* with the two-fold access array of satellites 1 and 2, [2Fs_2] to give the two-fold access array of satellites 1, 2, and 3, [2Fs_3]. This process of unions and intersections continues down each column proceeding from left to right until element (4,24) is finally calculated, [24Fs_{24}], which is the four-fold access interval array for the entire 24-satellite constellation. It is from this set of rise/set times that the coverage metrics are calculated. As a reminder, the PDOP constraint and minimum elevation angle are incorporated into the oscillatory access function, so these

constraints have already been applied to the intervals that the multi-fold access algorithm receives.

Table F-2: Multi-fold access interval algorithm

Folds	Satellite Identifier, s 1	2	3	…	24
1	$[1Fs_1]=[s_1]$	$[1Fs_2]=[s_2]\cup[1Fs_1]$	$[1Fs_3]=[s_3]\cup[1Fs_2]$	…	$[1Fs_{24}]=[s_{24}]\cup[1Fs_{23}]$
2	[0 0]	$[2Fs_2]=[s_2]\cap[1Fs_1]$	$[2Fs_3]=([1Fs_2]\cap[s_3])\cup[2Fs_2]$	…	$[2Fs_{24}]=([1Fs_{23}]\cap[s_{24}])\cup[2Fs_{23}]$
3	[0 0]	[0 0]	$[3Fs_3]=[s_3]\cap[2Fs_2]$	…	$[3Fs_{24}]=([2Fs_{23}]\cap[s_{24}])\cup[3Fs_{23}]$
4	[0 0]	[0 0]	[0 0]	…	$[4Fs_{24}]=([3Fs_{23}]\cap[s_{24}])\cup[4Fs_{23}]$

Appendix G

Results for Case-1b, Case-2a, and Case-2b

Case-1b, Case-2a, and Case-2b imposed the constraint that all satellites maneuvered would share a common *K* decision variable. The common *K* allowed for observations to be made regarding chromosome length and problem complexity but has little practical use since there is generally no reason to require such a constraint. As such, the hypervolumes and decision variable plots for these Cases are provided here for completeness but they have little value in the decision making process illustrated in Chapter 5.

G.1 One Satellite Lost, Idealized Assumptions

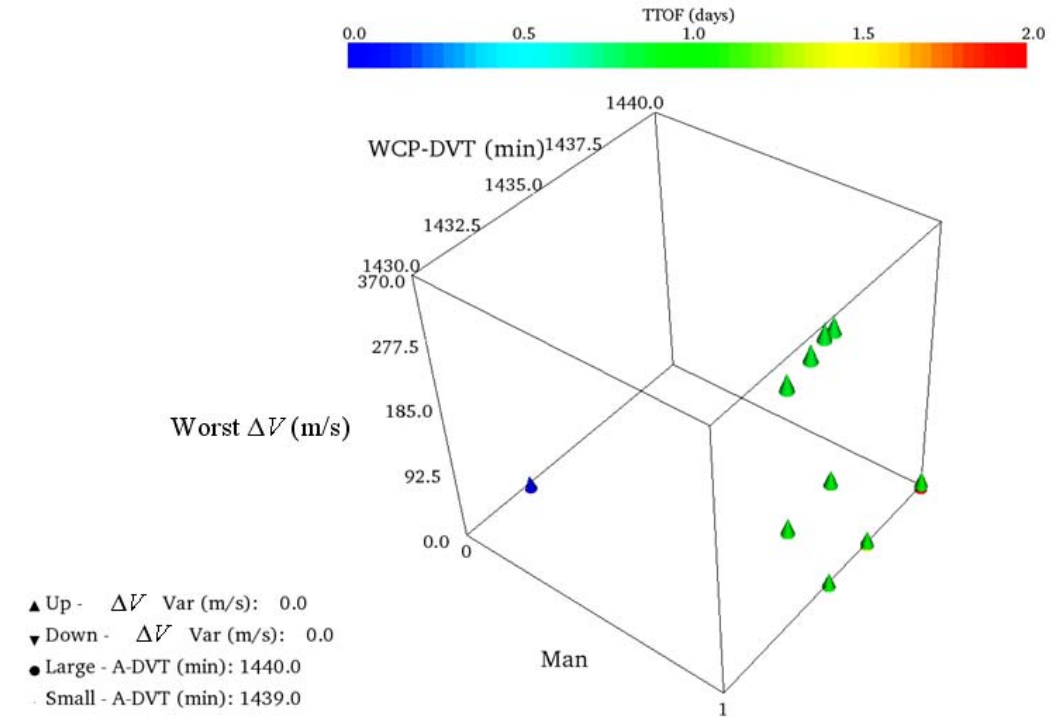


Figure G-1: Glyph hypervolume at run termination, Case-1b-ideal (one satellite lost, PDOP<2)

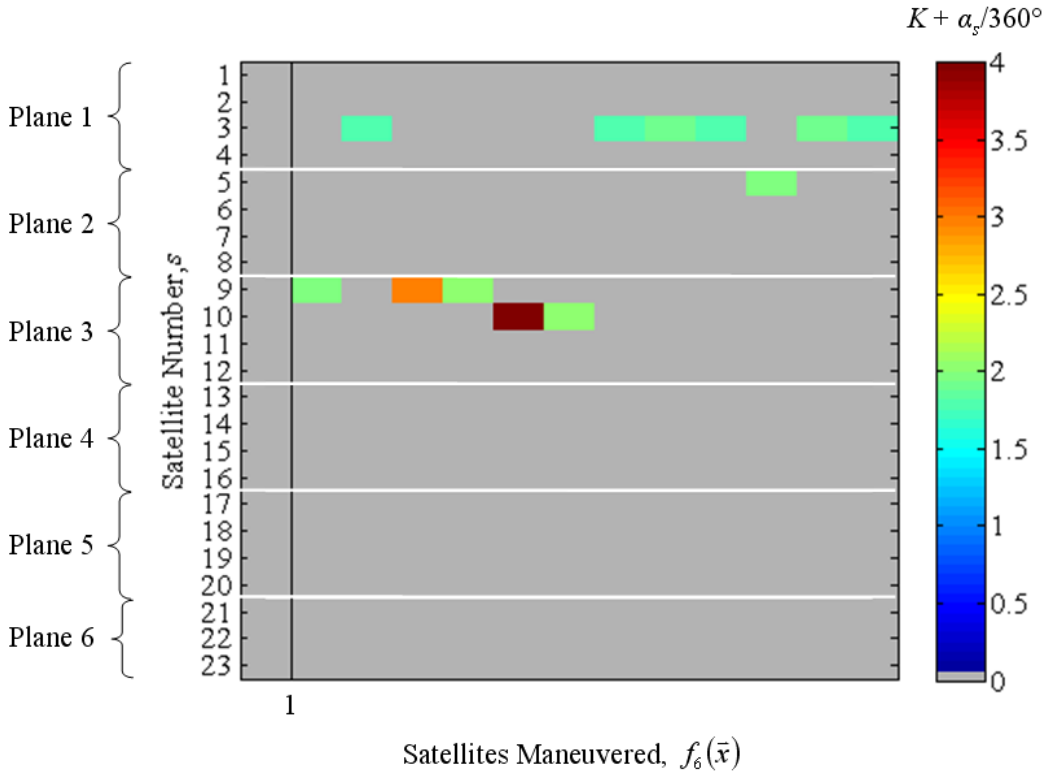


Figure G-2: Hypervolume decision variable plot, Case-1b-ideal (one satellite lost, PDOP<2) common phasing orbits, K, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\vec{x})$ ascending

G.2 One Satellite Lost, Operational Assumptions

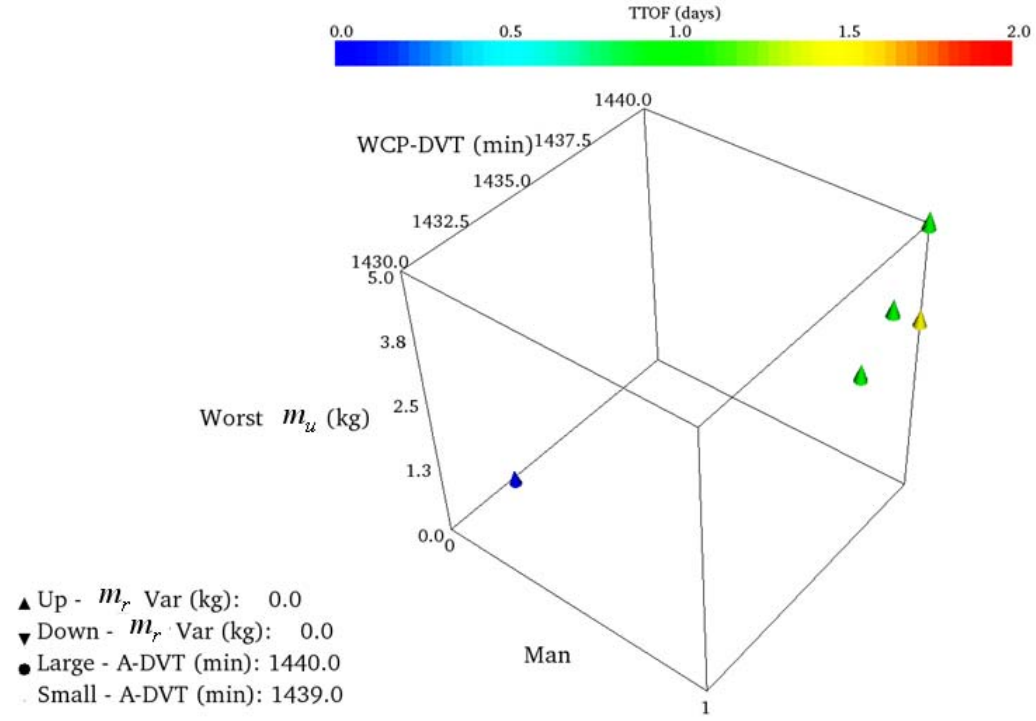


Figure G-3: Glyph hypervolume at run termination, Case-1b-op (one satellite lost, PDOP<2)

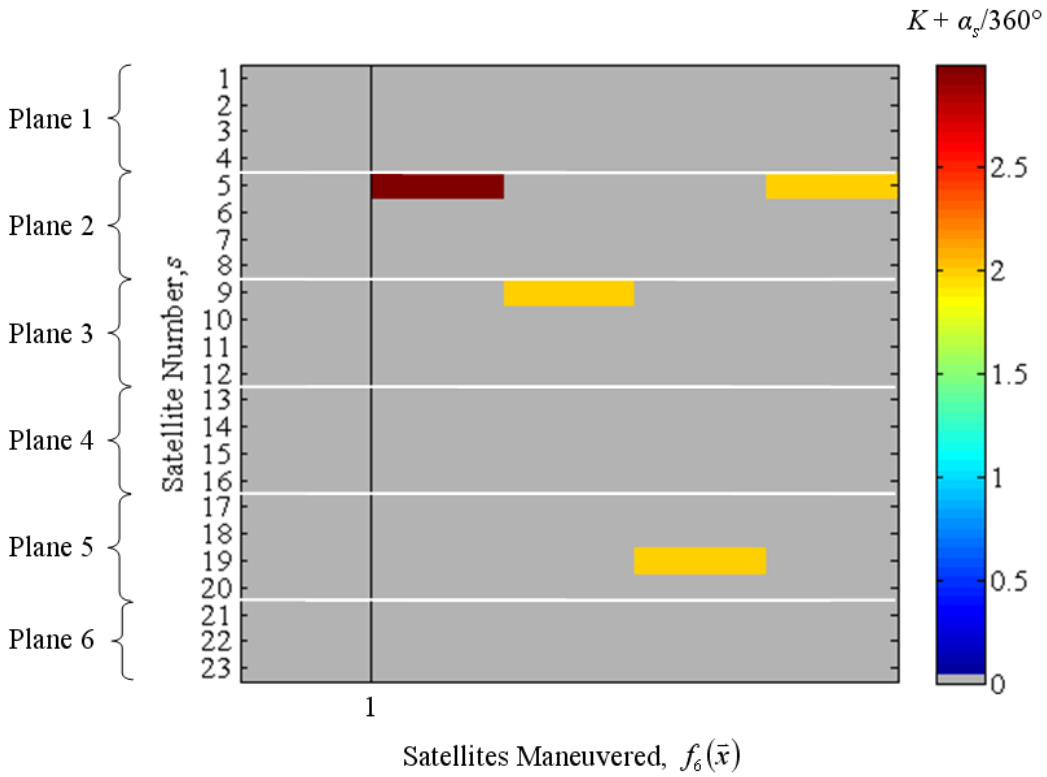


Figure G-4: Hypervolume decision variable plot, Case-1b-op (one satellite lost, PDOP<2) common phasing orbits, K, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\vec{x})$ ascending

G.3 One Plane Lost, Idealized Assumptions

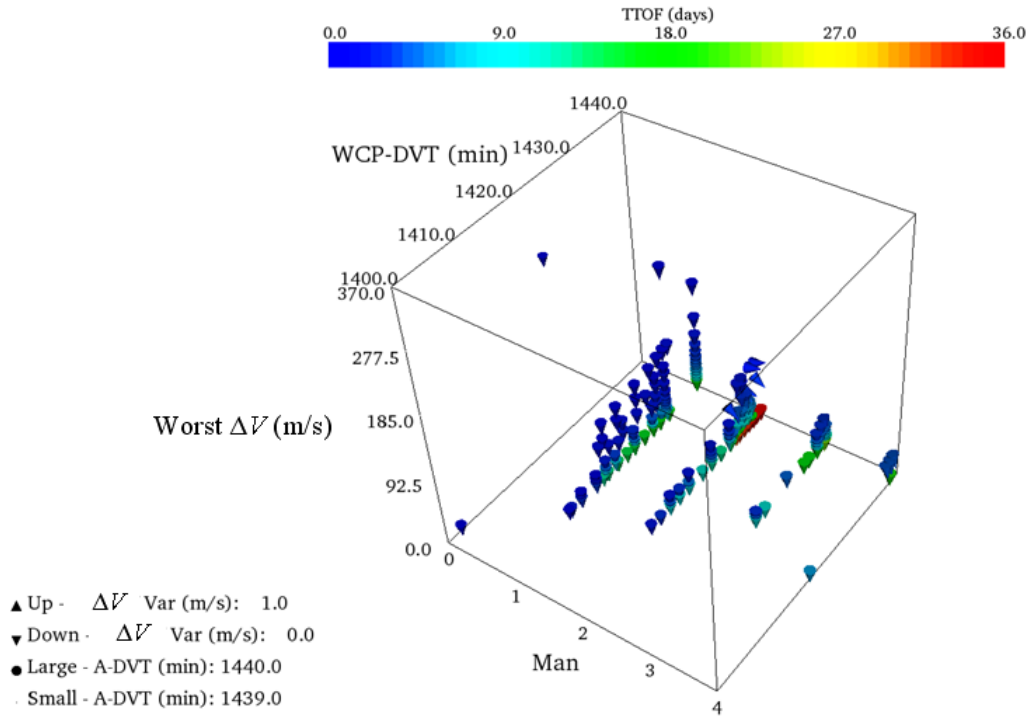


Figure G-5: Glyph hypervolume at run termination, Case-2a-ideal (one plane lost, PDOP<6)

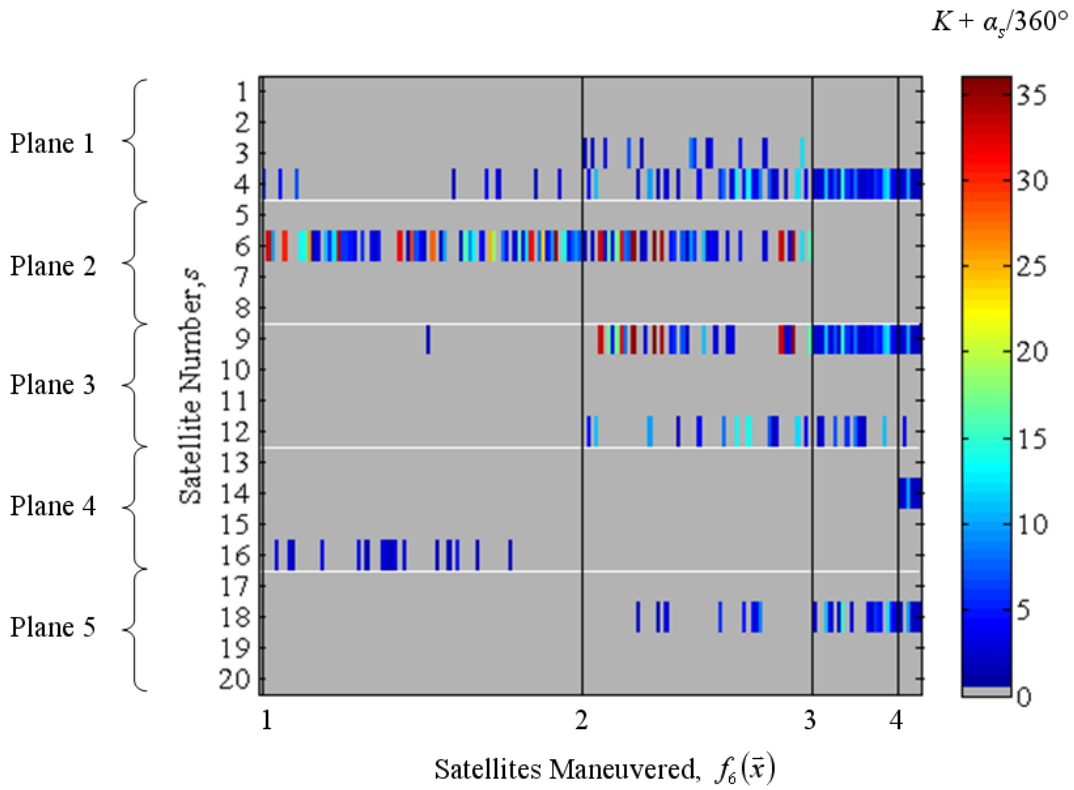


Figure G-6: Hypervolume decision variable plot, Case-2a-ideal (one plane lost, PDOP<6) common phasing orbits, K, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\bar{x})$ ascending

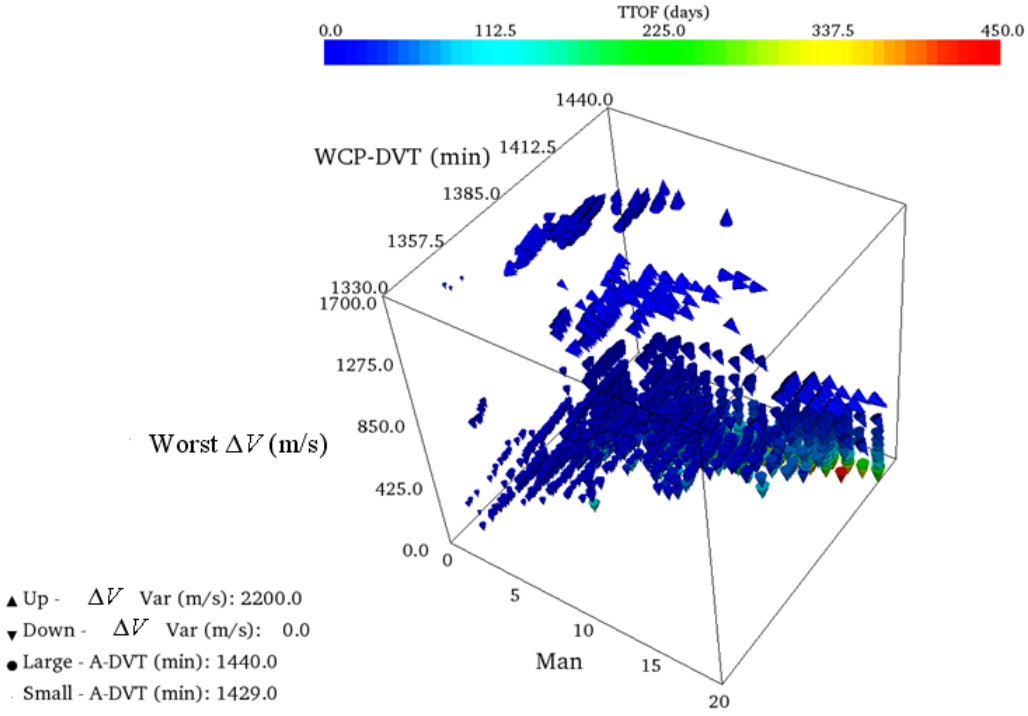


Figure G-7: Glyph hypervolume at run termination, Case-2b-ideal (one plane lost, PDOP<2)

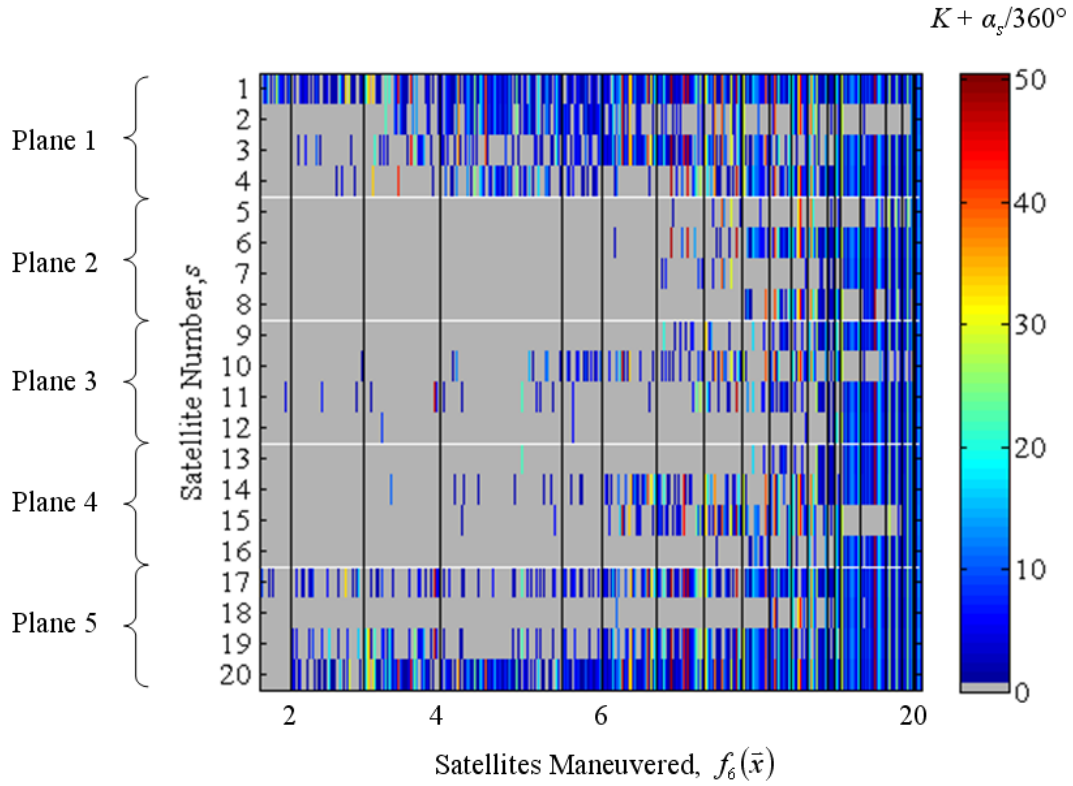


Figure G-8: Hypervolume decision variable plot, Case-2b-ideal (one plane lost, PDOP<2) common phasing orbits, K, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\vec{x})$ ascending

G.4 One Plane Lost, Operational Assumptions

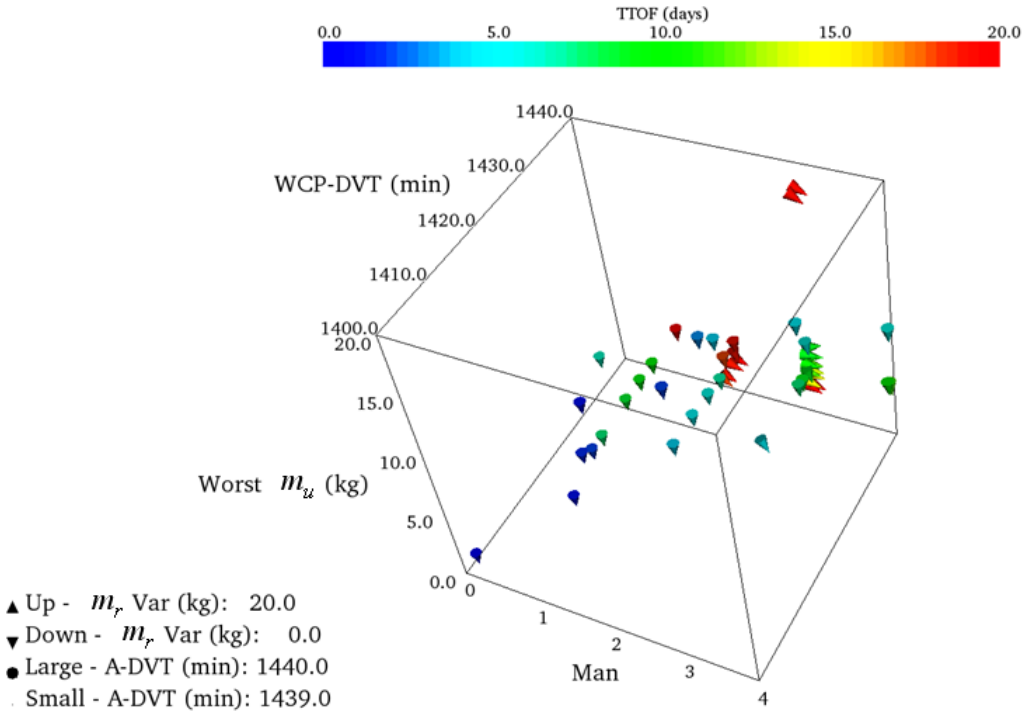


Figure G-9: Glyph hypervolume at run termination, Case-2a-op (one plane lost, PDOP<6)

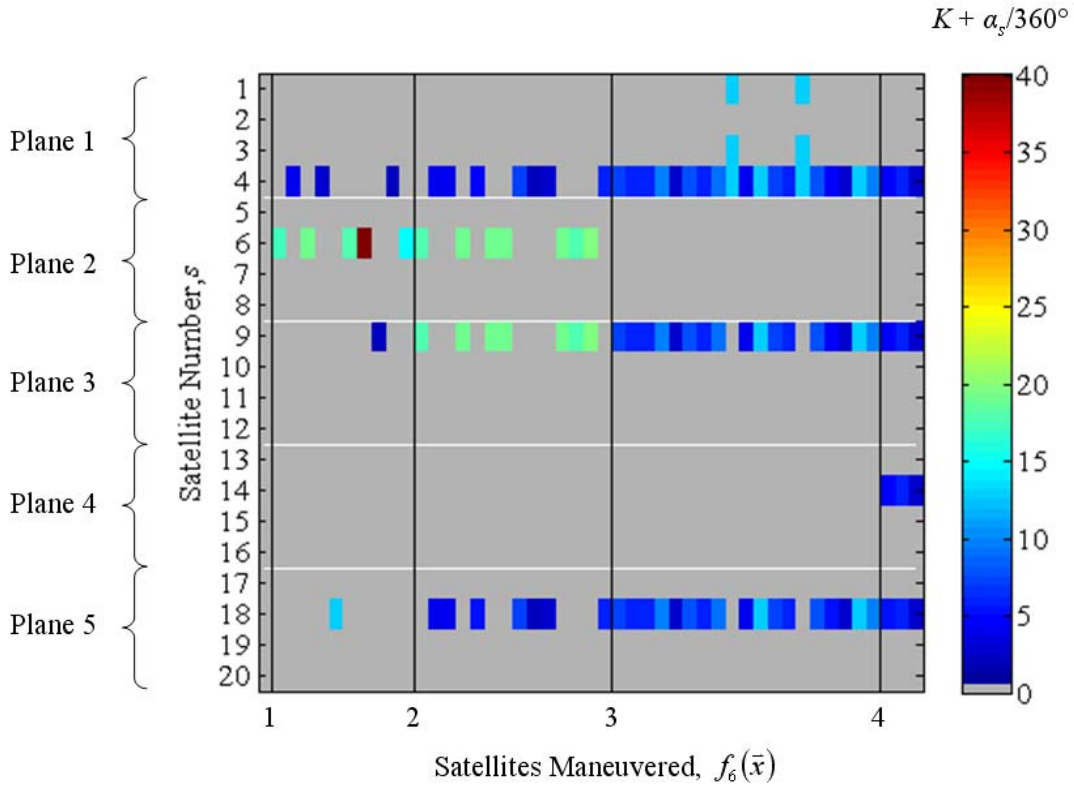


Figure G-10: Hypervolume decision variable plot, Case-2a-op (one plane lost, PDOP<6) common phasing orbits, K, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\bar{x})$ ascending

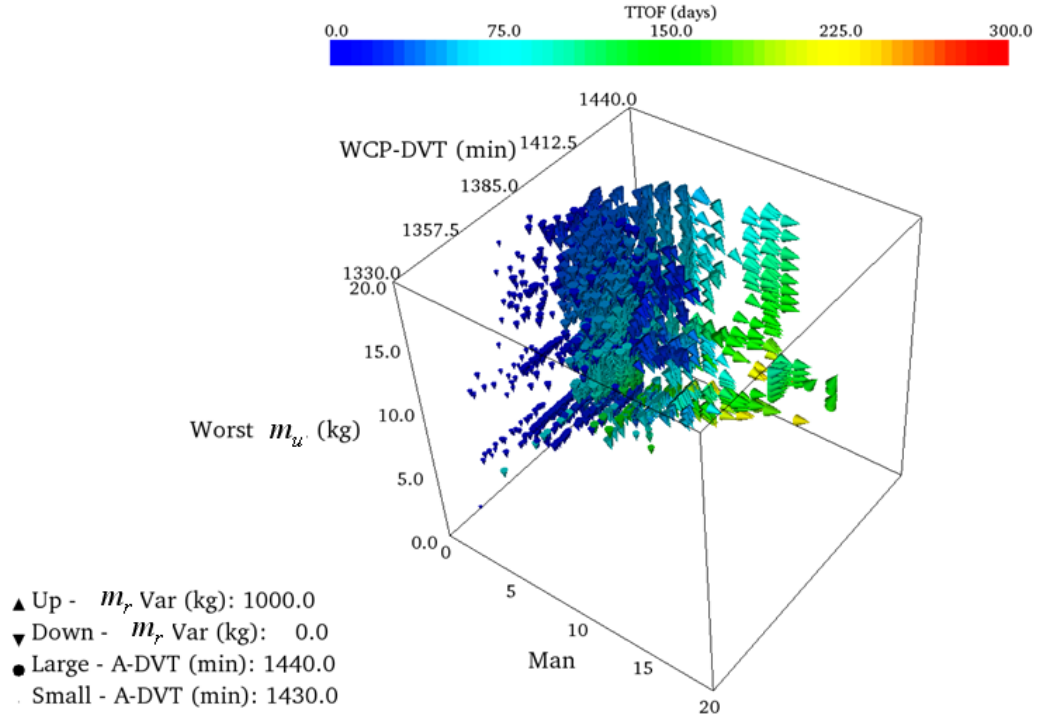


Figure G-11: Glyph hypervolume at run termination, Case-2b-op (one plane lost, PDOP<2)

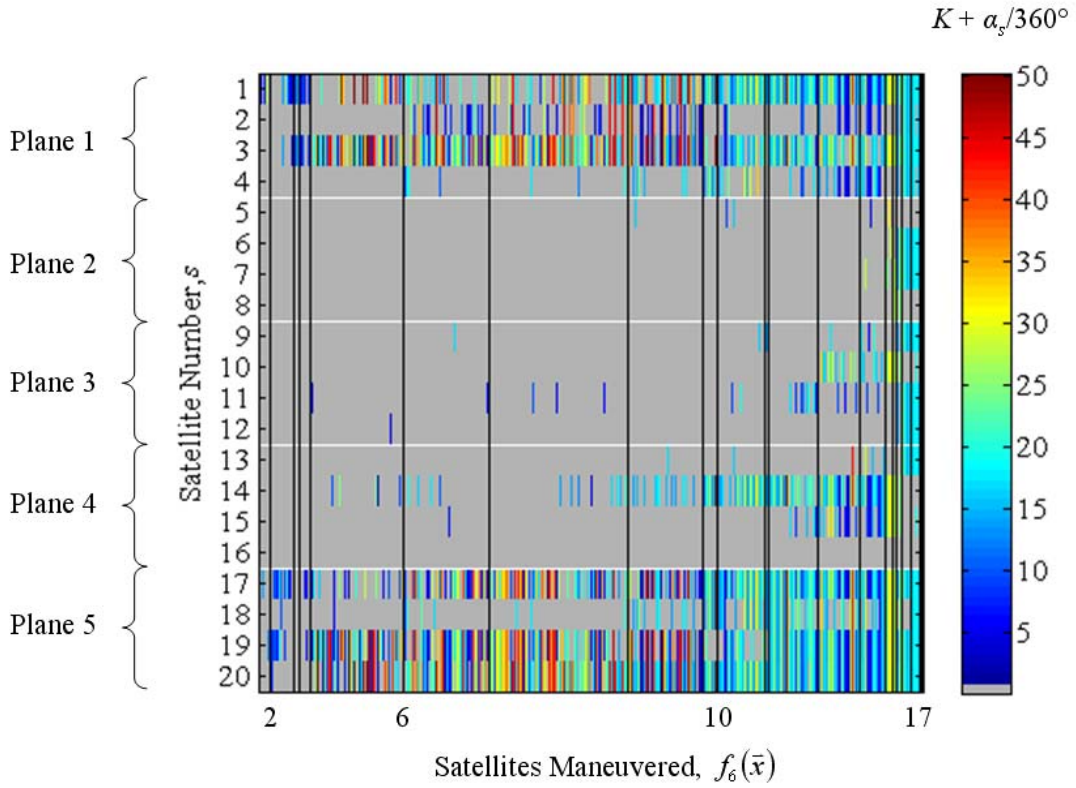


Figure G-12: Hypervolume decision variable plot, Case-2b-op (one plane lost, PDOP<2) common phasing orbits, K, plus change in phase, α_s, for each epsilon non-dominated design, sorted by $f_6(\bar{x})$ ascending

VITA

Matthew Phillip Ferringer was born in State College, Pennsylvania, on September 1, 1980. In 2002 he received his Bachelor of Science degree in aerospace engineering from The Pennsylvania State University (PSU). During his undergraduate years, he completed a cooperative education program rotating through Lockheed Martin Management and Data Systems, Boeing Integrated Defense Systems, and Boeing Space and Intelligence Systems. Shortly after receiving his Bachelor's degree, Matthew accepted a position with The Aerospace Corporation in Chantilly Virginia where he is currently employed as a Senior Member of the Technical Staff. While working for Aerospace, Matthew earned his Master of Science degree in aerospace engineering and a graduate minor in engineering mechanics from PSU in 2005. He subsequently enrolled in the doctoral program to study satellite constellation design and, specifically, optimization problems in that domain.

Matthew has served as President of Kappa Theta Epsilon, Cooperative Education Honors Society and as Secretary of Sigma Gamma Tau, Aerospace Honors Society. He is a Professional Member of the American Institute of Aeronautics and Astronautics (AIAA) and was recently nominated for the AIAA Young Engineer Scientist Award. Matthew is the recipient of the PSU Aerospace Pioneers Scholarship, the Honeywell International Engineering Scholarship, The Aerospace Corporation Graduate Fellowship, and The Aerospace Corporation's Inventors Award. Matthew's fields of interest include satellite constellation design, multi-objective stochastic optimization, evolutionary computation, decision support, and machine learning.