Space Systems Architecture for Resource Utilization

Space Systems Architecture for Resource Utilization:

A Workbook for Practitioners

By

Peter J. Schubert

Cambridge Scholars Publishing

Space Systems Architecture for Resource Utilization:
A Workbook for Practitioners

By Peter J. Schubert

This book first published 2021

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-6765-6
ISBN (13): 978-1-5275-6765-8

Table of Contents

Chapter 1 1
The Origin, Purpose, and Use of This Book

Chapter 2 8
Review of Scientific Literature and Resources

Chapter 3 54
Building a Resource Utilization Arc

Chapter 4 87
Techno-Economic Analysis

Chapter 5 129
Link to Other Arcs

Chapter 6 167
Outreach and Application

Chapter 7 192
Pathmaking and Collaboration

Chapter 8 219
Continuity and Growth

Chapter 9 237
Example Resource Utilization Arcs

Chapter 10 262
Appendices

References and Citations 274

CHAPTER 1

THE ORIGIN, PURPOSE, AND USE OF THIS BOOK

Abstract

This book guides the development of resource utilization arcs, from raw materials to delivered finished goods, and then weaves them into a complete systems architecture for the development of a Solar System economy. From broad concepts to specific technologies and their integration, this book provides a comprehensive approach to human enterprise beyond low Earth orbit. Intended as a working textbook, the user is guided through the practices of systems engineering and techno-economic analysis to prepare business plans for the comprehensive development of in situ resource utilization (ISRU) at a commercial level. This guidance identifies new areas for much-needed research and establishes the continuity of the team effort across the years to support ever-increasing sophistication and value for space-based manufacturing and services. This is the first book to address the pathway through which to achieve vital and sustainable practices in outer space. For example, the extraction of platinum from asteroids and the manufacture of solar power satellites from lunar materials together offer a hydrogen economy on Earth where all power and energy storage are derived from sunlight and water. The methods and concepts are sufficiently broad to appeal to a wide range of academic, industrial, and government efforts. A key goal of this work is to foster successful international collaboration to quickly bring the benefits of space utilization to humankind everywhere.

A. Origin

Space resources, such as vacuum, sunlight, minerals, and volatile substances, can be used to expand the human economic sphere, and provide many broad benefits and capabilities. Our home planet holds resources from space: valuable elements, which fell from the sky long ago, and sunlight, which arrives every day. In this **workbook**, a space resource is defined as something found beyond the Earth's atmosphere. We deal with **architecture** in the expanded sense of an integrated infrastructure for a **system of systems**, which accomplishes a goal or provides a valuable service. **In situ** (Latin for "in place" or "where it is found") **resource utilization** (ISRU) is the term for the local harvesting or processing of solids, liquids, gases, plasmas, gravity gradients, heat, pressure, or electromagnetic energy to meet a need.

At first, ISRU machinery and equipment will be built on Earth and transported to these resources. Over time, as the architecture becomes more complex and comprehensive, it will become possible to build ISRU factories from space resources. The purpose of this book is to guide the comprehensive architecture for each stage of this evolution towards the self-sufficiency of space settlements plus commerce and trade with the Earth and other planetary bodies.

The totality of skills and considerations involved in ISRU calls for a thorough treatment. The origin of this book is a desire to bring together a complete perspective for aspiring systems engineers and space architects. The chapters are designed to be used independently by those with focused interests, within a sequential study from start to finish, which is applicable to a wide variety of technologies and configurations.

A motivation for preparing this book is the dearth of such treatments, and the fragmentary nature of ISRU studies in an age where space activities are a low priority in many nations. Recently, there has been an increasing commercial interest in space. The existing half-century of space development has focused almost entirely on science, as well as on orbital communications or observation. Several start-up companies are pursuing prospecting, mining, propellant depots, and tourism. Many established aerospace corporations have long-standing programs in ISRU, awaiting a breakout. Private space industry, which is often funded by high net worth individuals, offers the possibility of rapid deployment of ISRU in a manner that is self-stoking. Those who learn from this book may be better prepared to capitalize on such possibilities.

B. Purpose

Humankind utilizes most of the Earth's arable and habitable land for food production and living space. The hydrological cycle is greatly affected by the activities of civilization, including the purity and availability of water. Subsurface water aquifers, which were eons in the making, are being depleted. Water tables are dropping all across the world. The sea level is rising due to warmer oceans and increased melt runoff of glaciers and polar ice caps. Higher sea level causes brine encroachment into wells nearby the shore, adding to water stress in many regions, especially in island nations.

Atmospheric gases are significantly affected by industrialization and the consumptive nature of affluent society. Isotopic analysis of the increase in carbon dioxide conclusively shows that this increase comes from the combustion of fossilized carbon in the form of coal, oil, and peat. Atmospheric methane has increased due to oil and gas exploitation, and also from agriculture, particularly from the belches and flatulence of ruminants raised for meat and milk. Methane is more potent than carbon dioxide in trapping long-wavelength heat radiation from the sun-warmed Earth inside the biosphere. The increased global warming contributes to the release of further methane from thawing permafrost in peri-arctic regions. Many other effluents of modern lifestyles further contribute to the greenhouse effect.

Humans who learn of the profligate lifestyles of those more prosperous than themselves seem to desire similar affluence. One amazing form of technology is air conditioning, which is increasingly sought out by people, and operated more than would otherwise be the case if global temperatures were not rising. More air conditioners mean more refrigerants, some of which produce so-called greenhouse gases (GHG), and these machines require power to move heat from indoors to the exterior air. More power means more or larger power plants. Centralized power generation and distribution from thermal power plants, fueled by coal, natural gas, oil, or uranium, requires cooling water to drive thermodynamic cycles to perform work on electrical dynamos. These fuels produce waste and, when the weather is too warm to cool the thermal cycles, the reactors are sometimes forced to shut down. Multiple, overlapping self-stoking cycles are accelerating these unsustainable practices, thereby hastening the day of reckoning.

Renewable energy, such as wind and solar are often cited as solutions, especially because the levelized cost of energy (LCOE) averaged over the entire life of the asset is lower than for single-use fuels. However, the intermittency of sun and wind require energy storage to meet the profile of electricity demand. Low-cost storage, such as pumped storage hydro (PSH) or compressed air energy (CAE) storage, are good choices, but are often geographically limited. Batteries are expensive, dense, and resource-intensive. Hydrogen storage requires a fuel cell to convert back to electricity. An elegant solution is baseload (“always on”) power delivered wirelessly from orbiting power satellites in a concept called Space Solar

Power (SSP). Launch costs for SSP powersat assets must be factored into the economics, as the launch energy must be considered in the metric, Energy Returned On Energy Invested (EROEI).

If **space resources** can be used to build SSP powersats, the environmental burdens of manufacturing and launching are not borne by the Earth's biosphere. SSP could be the ultimate answer to terrestrial power needs. With ISRU, minimal Earth resources are required, aside from early-stage factories which will be launched to initiate the process. SSP architectures provide continuous power regardless of weather, and with no regard to the Earth's day or night cycle. Wireless Power Transfer (WPT) from orbiting powersats is less than perfectly efficient, so there will be some degree of heat energy released into the biosphere but this will be at levels far lower per unit of energy relative to nearly every other form of power generation. The receiving antennas will be anchored; therefore, the necessary steel and concrete must be counted against SSP but, again, the ratio per unit of energy is superior to all others. SSP is available to nearly all population centers on Earth. So, while SSP is not perfect as an energy source, it offers so many improvements as to be worthy of vigorous pursuit.

With SSP powersats delivering the same level of power every hour of the day, they are likely to be sized such that excess is delivered at night. An attractive diversion load is the electrolysis of water to produce hydrogen gas. When it is used in fuel cells, which also draw oxygen from the atmosphere, the only effluent is water vapor. Fuel cells require a catalyst to recombine hydrogen and oxygen, and, at present, the platinum group metals (PGM) used for this purpose are drawn from ore bodies that were delivered to the Earth from meteorites. When ISRU methods are applied to platinum-abundant asteroids, this precious metal can be brought back to Earth so that, as the architecture matures, costs will be brought down, thereby making fuel cells less expensive. These two ISRU architectures, SSP and PGM retrieval, present humankind with the potential to supply all of their energy renewably from sunlight and water.

The purpose of this book is to provide self-sufficiency to outer space settlements and operations but it also aims to go further and provide benefits to all of humankind. Through the engines of commerce, ISRU entrepreneurs can earn their fortunes while providing lower-cost energy to the vast majority of people, as well as dramatically reducing the environmental impact on the Earth. Historically, waste from extractive practices has left a heavy burden upon the land. Space resources can be used to convert these materials into higher-value products, and to help close the cycles of consumption, which would lessen the pressure to extract more resources from under the Earth's surface. By integrating space resources into humankind's economic sphere, it may be possible to live more sustainably everywhere.

C. Use

Space resources and the environments in which they are found present challenges that differ from situations in which terrestrial practitioners have lived, worked, and studied. A direct application of historical methods is often possible, but may rely on the Earth's gravity field or on consumables that are only available within a mature economic system with complex and well-established lines of supply. Working in space, or on the surfaces of other worlds, invites new approaches. A combination of creative problem-solving, as well as the application of scientific rigor and engineering thoroughness, can often identify surprising new possibilities. And, while pointing out solutions to specific problems is certainly of value, this book aims to integrate multiple solutions where the overall architecture provides optimal benefits with minimal cost.

The practices of engineering optimization apply methods of iterative improvement to solution spaces where the number of degrees of freedom or the dimension of independent variables is sufficiently large that an unaided human brain cannot readily grasp them. Aided by non-linear regression, values of N around 6 can still be understood reasonably well. With many architectures having dimensions of N=100 or more, an astronomically large number of potential solutions exist. When seeking an engineering optimum, it is often useful to define a single performance metric, which is called a fitness function. A typical fitness function may be the weighted average of several individual metrics, such as the cost, risk, and duration of a project. Engineering optimization methods also exist where two or more fitness functions can be explored, leading to a contour plot where tradeoffs between the fitness functions can be found. At a large N, one can visualize the vast search space as having many local optima, some of which are nearly identical in their fitness function, but with very different constituent elements. The N-dimensional distance between these can be explored for high values to produce several solutions which are nearly equal but maximally different. Using this book's methods, it is possible for every practitioner to identify a different answer, each of which is very nearly as good as the other.

Each time the methods of this book are applied and an arc, or architecture, is defined to utilize space resources, these solutions should be archived and made available to other practitioners. It is through the successive improvement of each generation of problem-solving that new intellectual property is created. Successive studies, each with a unique and individual perspective, may uncover, reveal, or discover new avenues which previously lay hidden or unexplored. With the energy and hardware cost associated with getting to space, the field of ISRU will always require better answers. Practitioners of this book may pursue these solutions while working for a national space agency, through a private company, or perhaps a consortium of many distinct entities. When the costs are sufficiently reduced, and the benefits sufficiently improved, a stronger case can be made for the initial investment required to implement the arc. As subsequent studies find ever-better arcs, they will stimulate competition and spur on the increasingly rapid development of outer space resources.

D. Audience

The structure of this book lends itself to a full treatment over one academic year. It can also be sectioned so that it can be used by multi-disciplinary commercial teams over a shorter period. Task-oriented work teams in larger corporations and specific space technology courses may employ particular chapters effectively. It is also possible for distributed, focused teams to work together through the arc formulation by tying together the output from each group.

The practitioner is expected to have a basic understanding of physics, math, and engineering, as well as to have completed two years of college in such fields at a minimum. Practitioners with a lower level of preparatory background should expect slower progress, and a greater need to gather data and experience from mentors, or by conducting scholarly research (see the general reference list in Chapter 2).

E. Goals

The resources found in space have the potential to dramatically improve life on Earth, as well as to provide for large populations living remotely from our home planet. It has been argued that advances in technology make it increasingly easy for a deranged individual to spread harm to ever-greater numbers of people. The result of such logic is that the possibility of a man-made extinction event will increase inexorably to 100%. Bolides from space likewise pose an existential threat to modern civilization, as do other cosmic events, such as nearby supernovae. The upshot is that, for humanity to survive in perpetuity, it is imperative to establish self-sufficient settlements in multiple, independent locations. Couple this with the innate urge to explore and expand, and the motive to exploit the resources of outer space becomes an essential part of the human spirit, as well as the next logical step in our development.

CHAPTER 2

REVIEW OF SCIENTIFIC LITERATURE AND RESOURCES

Abstract

Vast bodies of work exist in the fields of space resources, mission architectures, systems engineering, techno-economic analysis, and design methodology. This chapter is a review of each area and a presentation of citations for further in-depth study. Exploration of these resources is strongly encouraged and should be divided among team members for efficiency. After the publication of this book, additional discoveries will occur and new approaches will be invented and developed. This chapter is not exhaustive, but sufficiently representative that the practitioner can learn keywords and search terms, with enough examples of where to find resources to form a foundation for additional research. A further aim of this chapter is to emphasize the importance of documenting, archiving, linking, and publishing the advances made by the practitioner's team for the benefit of others. The motivation to retain secrets for competitive advantage can be weighed against the good of society in choosing when, how, and how much to share. The patent is one such balance, where, in return for disclosing the means or configuration of a novel and significant invention, the patent assignee is granted a fixed, but lengthy period of monopoly. This allows the originators to benefit commercially from the invention, in order to be compensated for the effort in its creation, while also publishing and preserving the public record of how advances can be accomplished by one of ordinary skill in the art. Publishing and attending conferences is strongly encouraged as such participation can spark new ideas, introduce new connections, and provide a forum for the refinement of a burgeoning systems engineer's thought process.

A. Space Resources

Prospecting for space resources can be done directly by visiting, but this is expensive and time-consuming. Information can be obtained remotely using active or passive long-range sensors or gathered from space rocks that fall to the Earth. Antarctic hikers frequently see small, dark-colored meteorites lying exposed on top of the snow and ice. Many of these impactors have been traced back to their origins, such as the primordial nebula or disk, or ejected from the Moon or Mars following much larger impacts there long ago. Some of the dust which settles in one's home came from space. High-altitude aircraft sample these micron-scale fragments of meteors on specially-designed collection surfaces, which are only exposed in the stratosphere—high above any common dust kicked up from the Earth's surface and below the altitude where most shooting stars ablate into cinders [Fraundorf]. Of course, sample returns from the Moon and asteroids are the best source of information on space resources, where they can be tested using the full spectrum of analysis equipment available on Earth. The Apollo program included six manned trips to the lunar surface and returned 382 kg of various rock and dust specimens. Several sample-return missions to comets and asteroids have provided more profound insight into the formation of our solar system, and the nature of resources found on these smaller bodies orbiting the sun.

The Moon is remarkably similar to the Earth's crust in isotopic composition, which is strongly suggestive of a common origin. The Giant Impact hypothesis speculates that two planets collided 4.5 b.y. ago, spewing molten debris into a ring around the remaining body. This ring coalesced into the Moon, which began life much closer to the Earth, and is slowly (3.8 cm/y) moving farther away. While the isotopes and non-volatile elements of the Moon may match the Earth, their morphology is very different, consisting of layers of successively larger fragments of broken rock and dust called regolith ("blanket" + "rock" in Latin). The average particle size on the surface is 70 microns, about the thickness of a human hair. The most exciting news about the Moon began to emerge many years after Apollo ended, with the discovery of hydrogen near the lunar poles in "permanently shadowed

Lunar Orbit Recession Anomaly
The laser range finder left on the Moon by Apollo shows it is receding faster than expected. Sediment records, historical eclipse records, and computer simulations agree within 3%, but measurements are 20% greater! What is going on here? See [Riofrio] for one possible explanation.

regions" (PSR) where the sun never shines. Thought to be remnants of cometary impacts elsewhere on the Moon which spread out and then settled back due to gravity, these deep, icy crater bottoms never see direct sunlight because of the nearly perfect square of the lunar spin axis and the solar ecliptic.

Water ice has vast implications as a space resource. Humans on the International Space Station (ISS) use 12 liters of water per day, which is twenty times less than the average ground-dweller. Water is essential for good health and convenient for personal sanitation. When split into its constituent elements by electrolysis or thermolysis, the result is hydrogen and oxygen. Oxygen is essential for humans, who consume a half cubic meter (at STP: standard temperature and pressure) daily. Hydrogen has many uses: as a fuel, it makes an excellent reaction mass for a high-performance nuclear thermal rocket (NTR) and, when recombined with oxygen, it makes an excellent chemical rocket propellant.

Icy craters on the Moon are a harsh environment, being the coldest places in the Solar system. Similar frigid resources exist in PSRs on Mercury, where temperatures of about 40 Kelvin are typical. At such extreme cryogenic temperatures, our typical human intuition is a poor guide. Ultra-cold metals are brittle. Frozen regolith is as hard as concrete. Chilled mechanisms can seize. Silicon-based electronics, meaning most of the integrated circuits (ICs) ever built, do not operate reliably below about 230 Kelvin and would need to be heated. Batteries rely on ion conduction through electrolytes and permeable membranes and need even more warmth than ICs. Some new battery designs have the potential to operate down to 200 Kelvin, which is still far above the ambient temperature on the crater floor where the ice resides. The methods of heating the frozen ground include reflective mirrors to direct the sunlight where needed; wireless power transfer to in-crater assets; or surplus heat from a nuclear reactor, which is somehow delivered to where needed. Transferring heat energy is generally a short-range prospect, as any means of conveyance must be well-insulated to prevent loss. Such complex issues make ice harvesting in PSRs a rich opportunity for the practitioner of ISRU.

Mars probably has water below the surface, and certainly contains water in its polar regions. The most abundant resource of interest for space utilization is an atmosphere of carbon dioxide (CO_2). A chemical reaction discovered by French chemist P. Sabatier (for which he won the Nobel Prize) drives carbon dioxide plus hydrogen, in the presence of a catalyst and with sufficient temperature and pressure, to methane (CH_4) plus water. The water can be electrolyzed (split) into hydrogen and oxygen, and the hydrogen recycled to make more methane. The pure oxygen, which originally came from the CO_2, can be mixed and combusted with the methane to make a very nice chemical rocket engine [Lewis]. The Sabatier process can also be used to scrub carbon dioxide from the air inside a spacecraft and is used for this purpose onboard the ISS. The so-called "reverse water gas shift" reaction is a closely related avenue for converting CO_2 to CH_4. Furthermore, if water ice is available, then electrolysis can be used to separate into hydrogen and oxygen, which forms a high-performance fuel-plus-oxidizer combination for chemical rocket engines. While the H_2+O_2 rocket has a higher specific impulse (ratio of thrust produced per mass flow rate of fuel-plus-oxidizer, abbreviated "I_{SP}") than CH_4+O_2 by about 20 percent, liquifying hydrogen (for safe storage) is more energy-intensive and more prone to boil-off than liquid oxygen (LOX) or CH_4.

The Missing Planet
Before the predominance of the accretion model for planet formation, the Titius-Bode law led to wild speculation about the fate of Phaeton. What could overcome self-gravity and blow a planet into pieces? If our Moon came about because of two planets colliding, leaving Earth intact, perhaps Phaeton exploded on its own when the uranium in its core went critical. Did an ancient civilization on Mars split their rival planet for heavy metals using a planet-buster, thereby inadvertently sterilizing their own world and triggering the Earth's P-Tr extinction event (the "Great Dying") 252 m.y. ago?

Two valuable resource ore bodies associated with Mars are its two moons, Phobos and Deimos. Thought to possibly be captured asteroids, they may be somewhat representative of the larger collection of asteroids which lie in the Asteroid Belt between Mars and Jupiter. The Titius-Bode law is the phenomenological discovery of a geometrical progression of the integer harmonics of the planets' orbital radii and is accurate to within 5 percent out to Uranus; however, there is a gap between Mars and Jupiter. It was the search for this "missing planet," which led to the first discovery of an asteroid, Ceres, which is also the largest such body. Theories of planetary formation postulate that the massive gravity of Jupiter scrambled the components which might have formed into the planet "Phaeton." The mass of asteroids still extant is approximately one percent of the mass of a rocky planet and, while somewhat concentrated in a "belt," they actually occupy a wide range of orbits. Some asteroids fall to Earth as meteorites, aiding in their classification. Their fiery passage through the Earth's atmosphere of necessity changes their morphology and internal structure, such that carbon-rich asteroids (C-Class) are rarely found intact. Still-orbiting asteroids have been probed from the Earth using radio waves, such as the powerful radio telescope at Arecibo (Puerto Rico). Reflected sunlight can be analyzed by multi-spectral imaging and composition inferred by comparison to terrestrial minerals. Several asteroid missions by spacecraft have been conducted; the first few of which had limited success in returning samples to Earth for more detailed study.

Asteroids may have been the origin of some of the many moons found around the two gas giant planets: Jupiter and Saturn. Some of these moons may also have been comets, which have been captured by gravity on their fall sunward from the Kuiper Belt or the Oort Cloud. Kuiper Belt objects, including the former planet Pluto and its moon Charon, are generally believed to be agglomerations of ice and dust. The vaporization of volatiles when a comet is near the sun produces the tail. These vapors of water, and probably methane, ammonia, and possibly more complex hydrocarbons, become illuminated by sunlight, making them visible to the unaided eye on Earth. Some fraction of asteroids may contain a heart of volatile compounds that are protected from the sun by their rocky regolith shell with this outer crust being the condensed remnants after intermingled ices have boiled off. The possibility of water ice from a "dead" comet within the orbit of Jupiter could be a tremendous boon to space travel, in the form of fuel and life support materials. Large-scale space habitats with hundreds or thousands of people will benefit greatly from any nitrogen found in such bodies, as a nitrogen-rich atmosphere inside significantly reduces fire risk relative to a pure O_2 environment.

Harvesting gases from the upper reaches of the gas giants is an interesting concept. However, in the case of Jupiter, there are intense electromagnetic fields that accelerate ions in space, generating a constant storm of radiation. Radiation is bad for people and bad for electronics. These factors make gas harvesting even more dangerous than the basic requirement of a deep dive into a steep and unforgiving gravity well to scoop up fuel for fusion drives. Deuterium (hydrogen-2) can be found as heavy water wherever there is water ice. Helium-3 is an excellent fusion fuel when combined with deuterium because the flux of killer neutrons is much reduced relative to other fusion reactions. The ISRU practitioner might find ^3He escaping from the atmosphere of a gas giant, but another source is the flux of particles emitted from the messy fusion reactor of our home star, Sol. A perennial interest in lunar ISRU is harvesting ^3He from implanted "solar wind" particles on the surface. The price per weight for such fuel is astronomical.

The concentration of this precious isotope is so low that a great deal of excavation and heating of regolith is required to capture even minuscule quantities. Perhaps, when added to other processes for lunar regolith beneficiation, the harvesting of ^{3}He may become economical (see Chapter 9). Another, possibly richer source of this magic interstellar rocket fuel is the surface of Mercury, which receives about ten times the concentration of helium-3 from the solar wind relative to the Moon.

The Moons of the gas giants are especially interesting for ISRU development, as well as being of incredible scientific value. The ocean on Enceladus, lying underneath a thick shield of ice, yet kept liquid by the heat of the rocky core, is likely to contain ocean thermal vents such as those believed to give rise to the first forms of life on Earth. Another earthly parallel is Lake Vostok on the Antarctic continent, which has been buried for millions of years two kilometers below the surface of the ice cap and kept liquid by geothermal processes. Evidence of life there certainly demonstrates the potential for hardy organisms to be sustained in such extreme environments. The first penetration by man into this ancient body of polar water was criticized for failing to maintain perfect sterility, possibly contaminating the lake with an inoculation of modern-day organisms on the drill bit. Regardless of the commercial value of the water inside Enceladus, it must not be extracted without being preceded by careful scientific exploration and analysis. Scientific missions, such as the search for extraterrestrial life on the moons of Saturn, might be paying customers for the fruits of ISRU. Exploration and exploitation need not be at cross purposes!

Water Bear, Moss Piglet, Slow-Walker
One of the most resilient organisms ever is the Tardigrade, known by cute names, and being about the size of a sand grain. They have survived all 5 mass extinctions on Earth, and can withstand vacuum, radiation, dehydration, pressure, cold, and heat up to 270 C. Found *everywhere* on Earth, these 8-legged survivors are an inspiration for space explorers.

Listed in the reference section is a representative sample of scientific literature on space resource utilization. A review of book, chapter, article, and website titles will advise the practitioner of the long history of such studies and provide a solid foundation for more in-depth investigation.

Exercise: **Venus: Turning a Lemon into Lemonade**

Venus has a surface soaked by rain of sulfuric acid and is hot enough to melt tin (Sn) and lead (Pb). Above the corrosive and reflective cloud deck is an atmosphere of mostly CO_2 with some N_2. Gravity is similar to that on Earth and is closer to the sun by 28%. What resources might you extract from this hostile planet, and Earth's nearest neighbor?

Guidance: A fiction writer and NASA scientist describes using atmospheric carbon to fabricate transparent diamond windows that enclose floating cities [Landis]. What could these cities export?

Practice: Consider resources beyond matter, such as proximity to the Earth, proximity to the sun, and gravity, to envision services or sites which may provide value to an economy that spans the solar system.

SECTION CHECKLIST

- ☐ Not everything possible is economical
- ☐ Velocity change (Δv) is a huge cost driver
- ☐ "Time is money": distance diminishes value
- ☐ Seek out complementary resources
- ☐ Consider the economic value of services and sites
- ☐ Asteroids have many hidden secrets
- ☐ Fuel and life support are hardy markets
- ☐ Science can help pay the bills

B. Mission Architectures

Defining a space architecture is best approached by describing what it is not. Sorties in space are called a mission if they have a direct and easily-explained purpose, function, and objective. The original concept for the there-and-back Moon shot undertaken by America after the challenge laid down by President J.F. Kennedy was described as a "direct ascent." An enormous rocket would throw everything needed to land, take off from there, and return to Earth directly. An alternative advocated initially by Apollo leader W. von Braun was an Earth orbit rendezvous (EOR) involving a space station in low Earth orbit (LEO). This scheme involved multiple rocket launches to put the components into orbit, where they could be assembled, and provide a platform to bridge the Moon-bound Apollo astronauts between the Earth launch and lunar orbit insertion. The reconfiguration of the transfer vehicle to the Moon was changed close by to the Earth, with the potential for a relatively safe return if the rendezvous mating of spacecraft went awry. The EOR scheme would not only answer the Kennedy challenge but create a resource for further scientific studies, which may support a diverse array of future missions. This concept was perhaps the first recognizable “space architecture”.

The architecture eventually selected by NASA was called Lunar Orbit Rendezvous (LOR), which was strenuously advocated by J.C. Houbolt over the dominating paradigms of direct ascent and EOR. Houbolt’s concept involved greater risk, in that the rendezvous occurred around the Moon three days from the safe haven of Earth. At the time of this debate, orbital rendezvous had not yet been attempted and was widely viewed as unacceptable. It is a testament to perseverance and belief in a more elegant solution that Houbolt eventually appealed directly to the NASA Administrator, bypassing the chain of command and “going over the heads” of his superiors. This move might have become what is called “career suicide.” However, the LOR concept prevailed, as the size and number of Earth launches was significantly lower than EOR or direct ascent. With the accelerated time table established by Kennedy in 1961, LOR was perhaps the only way to meet the requirement. The orbiting command/service module, which awaited the return of the lunar descent/ascent vehicle, was available to perform other functions, so the LOR meets the definition of an architecture.

Until the Space Age, architecture referred to the design of functional structures and interior spaces for human use [Karlen]. In considering space resource utilization, an architecture may also be thought of as multi-purpose infrastructure. On Earth, infrastructure includes assets such as bridges, highway systems, rail systems, networks of gas pipelines, airports, sewer systems, a health care system, the Internet, cell phone networks, communications satellite fleets, and the global positioning system (GPS). Infrastructure is generally considered the foundation or backbone upon which human operations and commerce can manifest. The word itself is derived from Latin roots meaning “under” or “within” and “to build,” so a simplistic example in space resources might be a “cycler”—a vessel placed in a looping orbit to which smaller

Cyclers for Interplanetary Transport
First envisaged for repeat trips to Venus and Mars [Hollister 1969], the “architect and cyclist” E. “Buzz” Aldrin re-energized the Mars Cycler [Aldrin 1985], which was further developed by scientists at NASA’s Jet Propulsion Lab [Friedlander].

vessels can rendezvous to later detach closer to their destination. A broader definition, which has been adopted in this work, is that space infrastructure is an asset that performs multiple functions or serves multiple clients, and supports multiple missions. Space architecture is then the design of functional space infrastructure.

In the preceding list of infrastructure elements on Earth, the words "system" or "network" appear frequently. In space, the design of an architecture is sometimes considered to be the study of "system of systems" (SoS), whereby wholly independent and stand-alone functions, missions, hardware, or other space assets are brought together to interact and achieve some goal, purpose, or capability which would not be otherwise possible. Space lacks the simplicity of laying down iron rails which reliably connect one location with another, can carry anyone's train, and needs little further action or energy. All space assets are in motion. Space architecture is complex and dynamic. As evident from the options of direct ascent, EOR, and LOR for the Apollo mission, there are often many possible solutions. Determining the optimal architecture is, therefore, of prime concern to the purpose of this book.

What is optimal? One answer is low cost. Another answer is low risk. And another is short duration. Other factors might include intangibles, such as economic externalities and aesthetics, or perhaps national pride and job creation. Some architecture decisions are made by one person who may be the national leader, the financier, or the one with the best reputation or grandest charisma. More often, there is a committee consisting of diverse individuals, each bringing a different perspective. A chairperson or facilitator may guide such a committee towards a consensus decision based on which factors are considered most important. For organizations extolling democracy, a vote may decide the approach. When multiple choices exist, which is typical for space architectures, a more sophisticated approach can help. It is also worth noting that in any organization, there may be minority opinions that do not get full consideration, as almost happened with Houbolt and LOR. Furthermore, there is always the possibility that a "better" solution exists but no one on

Voting Schemes
Majority vote is simple and universally understood. Super-majority votes (e.g., 2/3 required) encourage greater debate to achieve consensus. When more than two choices exist, ranking may be used, or multi-round voting to winnow out a favorite. Many such schemes are available to try out.

the team has thought of it. How, then, to achieve an optimum architecture.

Presented here for the first time is an architecture optimization methodology that attempts to gather the best ideas from the most number of people and to discover architecture options that successively approach a balanced outcome decided in advance of the process. The first step is for the decision-making community, or authority, to determine the proportional weight or importance of each relevant factor. Each factor receives a coefficient, or pre-multiplier representing its contribution to the consideration of a final decision. These weights or coefficients are percentages, which sum to 100. Engineers prefer weights that add to unity (1.000) but, in the general case, such architecture decisions involve more than just the technical community and percentages are universally understood. Inherently numerical metrics are easy to use, such as money, time, and jobs created. Measuring risk is slightly more challenging as subjective judgments are all that are available for potential outcomes that have not yet occurred. A formal risk assessment method can guide a group to a Risk Priority Number (RPN) for each hazard or unwanted outcome. The RPN can then be sorted and ranked. The RPN is a product of three numbers, usually subjectively determined by a group, and each defined on a scale from 1 to 10. These are (1) Occurrence, or how likely a given unwanted outcome may be; (2) Severity, or how consequential would be the actuality of such an outcome; and (3) Detection, or how difficult would it be for a person or sensor to become aware of the hazard early enough to mitigate it. Higher numbers are worse, and so a RPN of 1000 is extremely bad. Assessing different architectures can be accomplished by comparison of their respective RPNs, either the highest, or the average of the top three, or some other formula (please see Chapter 5, Section B, for more on risk assessment).

For considerations such as aesthetics, reducing it to a numerical value is anathema to many. However, we will shortly explain how a computer algorithm can derive an optimal architecture so that a suitable method can be found. One answer is to present alternate architectures to the stakeholders and community and ask them to rank or vote for them. This is a common practice in terrestrial architecture and it is often heavily influenced by authority, reputation, money, or ego. Regardless, there must be some human assessment involved which is reduced to a number, perhaps called the Aesthetic Quotient (AQ). A team of artists or community leaders may complete a poll of their responses, with their combined scores for each alternative captured as a number between one and five stars. In the simplest case, the AQ may be 100 or 0 depending on the decision of the supreme authority involved. In any case, the RPN, AQ, cost, duration, job creation, and any other metric are combined into a single equation representing the "fitness" of any given architectural choice. In engineering optimization, the concept of a "fitness function" is central to the use of powerful tools to discover the highest possible value. An example of a fitness function is shown below:

$$FF = 40 * cost + 25 * duration + 25 * \left(\frac{1000 - RPN}{1000}\right) + 10 * AQ$$

For a computer algorithm to optimize an architecture, it is vital that the fitness function is determined *a priori*. This exercise focuses the decision-making community to agree up-front on the relative value of various elements when considering a "good" architecture. Such agreement will not be trivial, and will likely involve contentious behavior and frustration among the participants. The wise practitioner will facilitate an outcome through a transparent process to assure each interlocutor that their viewpoint has been heard, considered, and included. All involved should be aware of the significance of subjectivity throughout the process, even in assumptions behind the cost estimates. As space architectures may span decades and may involve tremendous sums of money, transparency of the fitness function process is absolutely required. No one can expect any such exercise to anticipate all possibilities or to satisfy every participant, but the archive of the process by which the fitness function was derived must be available for future challengers to review. Such an archive provides evidence that a "best effort" and "due diligence" were exerted to make the process as inclusive and careful as was possible at the time.

The second step in this new method of architecture optimization is to collect all the best ideas. For an architectural application such as Space Solar Power, there are at least six categories and possibly as many as 30 unique architectures found in the published literature [Potter]. In this step, every known proposal is included with the tacit assumption that even the most humble or the most confusing or the most easily-dismissed concept may contain elements that support an optimal architecture. Each proposal is broken down into elements or subsystems which, when taken together, support an architecture. These subsystems are formulated as mathematical functions. These functions have inputs, such as raw materials, labor, embedded value, energy requirements, time required, and whatever other things, services, or considerations must exist for that subsystem to perform its function. Outputs of the subsystem functional block include the role it plays in supporting the architecture, plus any waste or effluent produced by its function, including speed gained, value-added, time required, risk added, environmental impact, and more. The number of inputs to the functional subsystem block is called its arity. Each functional subsystem block thus created must have the same arity. Because all blocks have a common arity, many will have unused, or null, inputs. The reason for universal arity is that, once every architecture has been captured as a functional block, they can be mixed and matched to build an architecture. The fitness function serves as a guide so that the optimization algorithm can evolve ever-better potential solutions.

Imagine the subsystem/element blocks of each architecture separated or removed like peas from a pod, and poured into a soup pot. The soup is "stirred" to mix the blocks so that any unique connection between those blocks, which may have had a common origin, no longer exists. Any ladle, or "spoonful," of the soup will include a random mix of blocks, scrambled to remove human bias and to make every block available to every other one. Now, expand the analogy of the blocks from being simple peas to being magnetic peas having a plurality of north and south poles (inputs and outputs). When one single block is drawn from the "soup," it may emerge attached to one or more other blocks. These are aligned such that the output of one block is "magnetically" attracted to the input of another block. Visualize now that, as the two blocks are starting to be withdrawn, more blocks snap into place, again linking outputs with inputs. Continuing this approach, one can imagine creating a train or a "tree" of connected elements/blocks. Being functions, they will convert inputs into outputs all along the branches of the chain. The outcome may be an architecture; however, it is more probable that the chain/train drawn from a randomized pool will produce non-sensical results. The key concept is that there remains a non-zero chance of creating an architecture in this way. It is more likely that, among these random groupings, there are subsets of a "good" architecture (as measured by the fitness function) within these chains or branches of functions. The next step introduces the amazing power of an engineering optimization tool called Genetic Programming.

Cyber-Sex

Natural selection by sexual reproduction between fit individuals adaptive to change is the driving force in the evolution of species. Computer algorithms using "evolutionary optimization" adapt this to produce ever-fitter solutions to engineering problems by mixing between pairs of better solutions in each optimization "generation."

Step three in architecture optimization is to use a modified method of Genetic Programming to assemble chains or trees of functional modules that build an architecture. The General Theory of Evolution inspired J. Holland to apply similar reasoning to optimizing functions having a large search space, a high number of input functions, non-continuous output functions and, possibly, many nearly-equal local optima. This method is called Genetic Algorithm optimization and it applies to existing functions that can be run by a computer [Holland]. In the case of space architecture, the function is not available in advance. Thus, it is a further refinement of GA called Genetic Programming (GP) by which a computer algorithm can create a computer algorithm. Using GP with automatically defined functions (ADF) is how useful subsets of a given chain/tree can be captured and reused in successive iterations [Koza] to optimize the fitness function.

The GP method begins with a set of basis functions analogous to our "magnetic peas," and pulls a population of chains (including branched chains) from the pot of pea soup. The population may contain several hundred, or perhaps several thousand such chains, each of which is evaluated using the fitness function. The first time this is performed, the results of the fitness function will generally be very low on our 0 to 100 scale. Of course, being random, some will have slightly higher scores than others. The evolutionary aspect of GP enters with the next operation, which is **sexual reproduction** among those original chains. Following the logic of the selection of the fittest, pairs of chains slated for reproduction are selected based on their probability-weighted fitness function scores. Among each individual chain, a specific link is randomly selected for fission. From two existing chains and by pair-wise swapping the portions before and after their respective fission site, two new chains are created. Because a GP chain will generally include branches, the fission point could occur at any given location, such that the two "offspring" may include one which is smaller than and another which is larger than the "parents."

GP Cleans Your Home
An example of how GP can evolve complex algorithms from simple mathematical components can be found in the development of a routine for a robotic vacuum cleaner to efficiently navigate and treat all surfaces within the home. While a pre-programmed search routine paired with obstacle sensors may work okay; in a more complex environment, a GP can generally evolve a superior algorithm which achieves higher performance by its optimization of a fitness function.

Generally, one of the offspring will have a higher fitness function than either of the parents. An additional element found in nature is **random mutation**. When creating a new population in this way, surprising new combinations might emerge, although many will fail. Random mutation follows the rules of arity and maintains the rules of constructing a valid chain. The number of such mutations is a small percentage, with their "mutant code" brought forward to the subsequent generation. A third process is to allow for selected high-scoring individuals to be brought forward without modification to a subsequent generation. This practice, which is called **elitism**, gives a higher probability of survival for those solutions which may have advanced far faster than the average. Upon creating a second generation via sexual reproduction (with replacement), mutation, and elitism, the new population is evaluated for fitness function scores. The process repeats with a third generation, stochastically selecting for sexual reproduction, mutation, and elitism. The stochastic nature of the process prevents stagnation. There is never a guarantee for the best solution to always win—although it is certainly given a greater opportunity. These GP operations generally require high performance computing (HPC) resources, as the complexity of the architectures can become quite significant.

The twist on GP, which was first introduced in 1999, is to insert an additional, separate optimization step in between each generation. It can be understood that a functional block from one architecture, when used in a very different architecture, might contain adjustable parameters which happen to be outside of a useful range. As a simple example, consider the payload capacity of an Earth-launched vehicle. Many architectures require a launch, and may require either a large number of small rockets or a small number of large rockets. The adjustable parameters, in this case, would be the fairing width, the payload mass, and the ultimate altitude the rocket can attain. Of course, these are interdependent, so the parameters are constrained by the realities of the rocket equation and existing material science. The new twist is to perform an optimization of these adjustable parameters at each iteration of the population. While GA can be a useful tool for this purpose, a more elegant approach is to use Particle Swarm Optimization (PSO) [Eberhart], which generally runs faster than a GA. By using the same fitness function, the burgeoning architectures at each generation of the GP process have the best chance to contribute their compositions to subsequent generations [Schubert 1999].

The overall result is that the highest score on the fitness function of the best individual architecture of the population will increase generation by generation. Following a classic learning curve, progress is rapid at first, followed by slower convergence towards a global optimum with subsequent generations. For problems having a large search space, the advances of fitness function by generation will asymptotically increase towards some ultimate “best” configuration; however, from a practical standpoint, once further increments become sufficiently small, the simulation is terminated. One individual that has the best overall fitness is then selected as the architecture of choice. In this way, the selection is accomplished by an algorithm that all can see and understand, guided by a fitness function which was decided in advance, so the outcome is as objective as it can be. The final solution is also likely to be superior to any single given architecture invented by a human.

Some people may feel that this “winner takes all” approach is stressful. Consider a population of thousands of candidates in the last stages of the learning curve, where the divergence between candidate architectures no longer changes as rapidly. There may be, among these nearly-equally-good candidates, great diversity in how they achieve their high value of the fitness function. D. Loughlin developed an approach that tests nearly-equal candidates for the diversity of their solutions. One method is via an assessment of the numerical “distance” between candidates. For a simple list of numbers, this can be the Hamming distance, although more complicated measures of diversity are undoubtedly possible. When applied to the GP+PSO optimization at the asymptotic stage of the learning curve, where improvements are marginal, this “multiply-generated alternatives” (MGA) approach can present a list of essentially equally-valued space architectures, which are as different as possible [Loughlin]. A reasonable number to present might be two: a primary solution and a backup. Or, one may present a portfolio of between three and five solutions, from which the final decision can be made based purely on human negotiations. At this stage, the objective preference for one over another is moot, so the propensity of humans to inject themselves and their personalities into a decision can be suffered without sacrificing performance.

Exercise: **Personal Fitness Function**

Develop a fitness function for your personal fitness. Include diet and exercise, plus sleep and stress reduction.

Guidance: Because you are the evaluation method, this process will be slow—practice patience.

Practice: Make yourself the experiment, and try various weightings to see which has the best results.

SECTION CHECKLIST

- ☐ Everybody has an architecture
- ☐ Draw the best from each idea
- ☐ Human decisions are rarely optimal
- ☐ Objective methods require transparency
- ☐ Include all stakeholders and the community
- ☐ Apply optimization whenever possible

C. Systems Engineering

Resources to learn more about systems engineering can be found, for example, in the NASA Systems Engineering Handbook (2007), posted on the nasa.gov website listed in the references section. Ryschkewitsch et al. summarize the collective wisdom of NASA's systems engineering greats in "The Art and Science of Systems Engineering." Fortescue addressed spacecraft systems engineering. Sheng's book, *Systems Engineering for Aerospace,* is concise and thorough, with many practical examples and tools to aid productivity. Sheng's focus is mostly on aircraft systems; however, the methodology generalizes well to space resources [NASA, Fortescue, Ryschkewitsch, Sheng].

Systems are made of components working together to perform some function, which would not be possible if they operated independently in the absence of interaction. Thus, a system is defined by relationships. In the field of Systems Engineering, complex and interdependent operations are managed between different disciplines, such as Mechanical Engineering, Electrical Engineering, Industrial Engineering, Chemical Engineering, Civil Engineering, and Biomedical Engineering. Mechatronics is an example of a portmanteau (a word mashup) describing an apparatus that combines mechanical and electrical functions. It may be a smart door lock function on a land vehicle which can be activated wirelessly via an encrypted signal

The First Systems Engineer (SE)
Systems thinking was first founded in the late 1950s by MIT's Professor Forrester, who applied this concept to testing social systems. In the 1960s, NASA dramatically expanded this concept to the hyper-complex world of rocketry and aerospace.

from a battery-powered remote control device and, upon verification, cycles a solenoid which rotates a cam that engages or disengages a latch between the door and the frame; in addition through a linkage, it actuates the interior manual door lock toggle switch. A Systems Engineer generally does not work at the granular level of components but studies the subsystems that must interact harmoniously to create the system function. The Systems Engineer is responsible for guiding the development of subsystems and their internal components so that the discipline-specific engineers are actively advancing towards a successful system operation. The demands on a Systems Engineer go beyond technical competence, as they require exemplary communication skills, mastery of a broad range of knowledge, and the persistent discipline of managing a project to a timeline, a budget, and performance metrics.

The first task of a Systems Engineer (SE) is to capture the "Voice of the Customer." The Requirements Specification is a document that is the heart of a contract between the Customer and the SE's organization. Financial and legal specialists are almost always involved, plus managers and the sales engineers who brought in the Customer. The SE should never show surprise (or annoyance) upon learning that the sales engineer has made promises to the Customer that are unrealistic or poorly-founded on the laws of physics and thermodynamics. Instead, a wise SE will work with the Customer to discover precisely what they need, separate this from what they would like to have, and present a logical and thoughtful series of alternatives and consequences (in terms of cost, duration, and risk) of the original dream. Many technically-savvy engineers would find this process frustrating and wasteful; they should, therefore, recognize and appreciate the value of the SE.