Space Systems Engineering

Fundamentals, Technologies, and Projects for the 21st Century

Table of Contents

Part 2: Transport Technologies

1. **Structural Methods** - Static, Dynamic
2. **Guns and Accelerators** - Mechanical, Artillery || **page 2:** Light Gas, Electric
3. **Combustion Engines** - Air Breathing, Internally Fueled (Rockets)
4. **Thermal Engines** - Electro-thermal, Photo-thermal, Nuclear-thermal
5. **Bulk Matter Engines** - Rotary, Coilgun, Railgun
6. **Ion and Plasma Engines** - Ion, Arcjet, Plasma
7. **High Energy Particle Engines** - Particle Rockets, External Particle Interaction
8. **Photon Engines** - Photon Sails, Photon Rockets, Photon Gun
9. **External Interaction Methods** - Magnetic, Gravity, Aerodynamic, Mechanical
10. **Theoretical Methods** - not currently supported by established physics
11. **Comparisons Among Methods** - Performance, Status, Cost

Part 3: Engineering Technologies

1. **Design Factors** - Technology, Availability, Physical, Integration, Human, Environment
2. **Subsystem Design**
3. **Resources** - Exploration Methods || **page 2** - Resource Inventory
4. **Resource Extraction** - Mining: Solid, Atmospheres, Liquids, Gas Giants, Particulates, Stars || Energy: Solar, Fission, Fusion, Orbital
5. **Processing and Fabrication** - Production Control, Handling and Storage, Materials Processing, Parts Production
6. **Assembly and Construction** - Assembly, Construction, Outfitting
7. **Verification and Test**
8. **Operation and Maintenance** - Operations Concepts, Tasks, Maintenance Concepts, Tasks
9. **Recycling Methods** - Waste Recycling, Closed Loop Life Support

Part 4: Projects and Programs

1. **Program Overview** - Goals, Summary, Structure
2. **Phase 0 - Research & Development** - Planning, Process, Sub-Phases & Tasks
3. **Phases 1 to 3 For Earth** - Locations: Starter, Network, Distributed, Industrial, Difficult, Extreme
4. **Phase 2B: Industrial Locations for Space** - Survey, Drivers, Concept, Details || **page 2** - Hypervelocity Launcher
5. **Phase 4A: Low Orbit Development** - Features, Projects || **page 2:** Orbital Assembly
6. **Phase 4B: High Orbit Development** - Features, Industries, Drivers, Projects, Integration, Details
7. **Phase 4C: Inner Interplanetary Development** - System Concept, High Orbit Skyhook, Inclination Station, Transfer Habitats
8. **Phases 4D-F:** - Main Belt & Trojan, Outer Interplanetary Scattered, Hills, & Oort
9. **Phase 5A: Lunar Development** - Features, Industry Survey Drivers, Projects || **page2** - Projects (cont.)
10. **Phase 5B: Mars Development** - Near Term: Phobos and Deimos, Mars Skyhooks, Mars Surface Systems || Long Term Development: Magnetosphere, Greenhouses, Full Atmosphere
11. **Phases 5C-E** - Venus & Mercury, Jupiter System, Outer Gas Giants
12. **Phase 6: Interstellar Development** - Nearby Interstellar, Nearby Exostellar, Farther Interstellar
13. **Orbital Mining** - Rationale, Mining Steps || **page 2** - Ore Types, Example Missions, Design
14. **Processing Factory** - Outputs, Process Research and Development, Factory Design
15. **Spaceport Network** - Concept, Applications, Design Parameters, Design Components, Design Issues

Part 5: Design Studies

1. **Conceptual Design for Human Expansion** - Goals and Benefits ||**page 2:** Conceptual Design Approach ||**page 3:** Requirements Analysis ||**page 4:** Evaluation Criteria ||**page 5:** Functional Analysis
2. **Environment Ranges**
3. **Seed Factories** - Conceptual designs for self-expanding automated factories (in separate wikibook).
4. **Open Source Space Program**

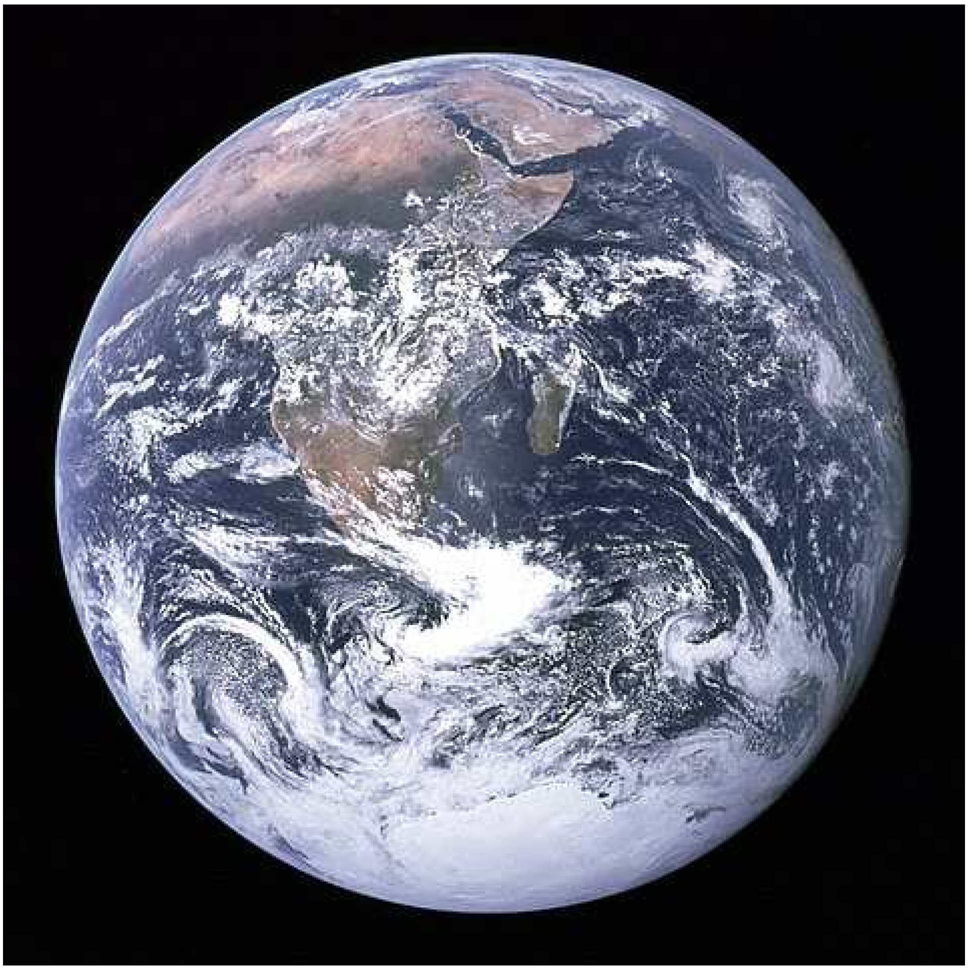
References and Sources

1. **Appendix 1: Fictional Methods**
2. **Appendix 2: Reference Data**

page traffic for this month.

Part 0: Introduction

Purpose of the Book


Earth as seen by Apollo 17.

As this photo of the Earth from a distance shows, our planet is located in space. So space is not a different place to go to: we are already there. We think of it as different because most of us spend most of our lives on about 1/8th of the solid-gas interface of one planet among the billions in the Galaxy. It is where we evolved, and where conditions are reasonably habitable for us, so we think of it as normal. In reality, our comfortable planet is the exception, and normal is unusable environments hostile to life as we know it. We can change that by technical means, and are starting to do so first by placing devices in space, then by creating habitable environments.

As of 2014 global space industry had already reached US$ 323 billion/year of economic activity. This is expected to greatly increase in the future as our civilization learns to use this environment and expand into it. For these and other reasons space systems engineering is a worthy field of study within the larger context of all sciences and engineering. Historically the main engineering challenges were transportation that could reach space at all, and the design of relatively small satellites with finite lives as cargo for such transportation. These topics are adequately covered by existing textbooks such as Sutton's **Rocket Propulsion Elements** and Griffen and French's **Space Vehicle Design**.

Starting with more recent projects, like the International Space Station (ISS), new design characteristics have become more important. These include permanent habitation, evolution and growth, sustainability, and industrial capacity. Future projects will require new types of propulsion, life support, production, and assembly. Single projects can no longer be considered in isolation, either. For example, the ISS depends on multiple launch vehicles to deliver crew, supplies, and hardware for expansion. This book is intended as an introduction to both the historical design challenges of transport and space hardware design, and the new engineering methods needed for future projects. The online format is also intended to address some of the limitations of traditional printed textbooks:

- Their high cost in recent years, which is addressed by making this an open-source Wikibook,
- The inability to include other media types, such as audio, video, simulations, and software.
- The relative difficulty of updating printed books in a fast moving field.
- The very large amount of information in this field, which is addressed by links and a supplementary library
- The lack of practice projects. Learning well involves more than just reading, it should include hands-on work. The latter part of the book includes real projects which interested readers or teams can work on.

Links to Additional Material

We cannot fit the entirety of a field of engineering into a book-length introduction such as this one. Therefore we will include extensive hypertext (World Wide Web) links to more detailed sources, such as the Wikipedia online encyclopedia. A **Rocket Science Library** is being set up to house some of the more useful publicly available articles and books to supplement the book. We cannot include copyrighted works without permission, but we can list useful ones and where to find them in both the **References** section of this book, and in notes in the Library folders. We are aware that popular culture incorrectly calls Space Systems Engineering "Rocket Science", as it is part of engineering rather than science and not exclusively about rockets. We will cheerfully adopt the popular label for our Library and hope that serious students understand the difference.

Engineers use other tools besides books to do their work, such as simulation software and spreadsheets. Wikibooks does not support including software files, so links to where these items can be found are intended to be added over time. A draft paper summarizing a real proposed program can be found at: **To Mars and Beyond**. This program will be used for examples within the book, and can be used by readers for practice in design.

Book Organization

The book is organized as a progression from science and engineering fundamentals, to subsystems, complete projects, and then combined systems which have linked projects. We will try to emphasize underlying principles and key design parameters, as those have more lasting value than, for example, the particulars of a current launch vehicle. We include many methods and concepts which are not in current use for several reasons. First, when starting a new project, it is a good idea to at least briefly consider **all** the possible alternatives, before narrowing down to the most relevant ones. Second, this is a future-oriented book. While a concept may not be useful today, knowing what the state of the art is and what technology areas to watch helps to tell when a concept will become useful. Third, having all the ideas in front of you might spur a new combination or application that had not been thought of before. This has actually happened to one of the authors (Eder) on more than one occasion, so I can attest to its usefulness.

Part 1: Science and Engineering Fundamentals

All of engineering depends on underlying knowledge from mathematics and the sciences. We therefore start with some prerequisites which the reader should be familiar with, and the more important principles, formulas, and methods which apply to space systems. The largest part of this is physics, including the motive forces and energy sources for space projects, but also some key ideas of the other sciences, including Astronomy, Planetary Science, and Chemistry. Following that we introduce general engineering methods, including the discipline of **Systems Engineering**. The goal of systems engineering is to optimize a complex system over its life cycle. Most space projects are complex enough that this discipline is very useful. Complexity demands methods of tracking the details to keep the project as a whole a coherent effort. We introduce the tools that engineers use in their work, followed by the variety of specialty engineering disciplines such as structural and electrical design which are used in parallel with the overall system optimization. Most projects require specialists in different areas, because there is too wide an area of knowledge for one person to cover all of it.

In early concept development one person might do all of the work on a project. As the design progresses, the amount of detail and knowledge required makes it more efficient to use teams of specialists. This in turn demands efficient coordination within a project team. The general topic of project economics addresses project organization and management, funding, and financial analysis. The last two sections of Part 1 consider existing projects and the areas for future projects. One needs to compare proposed future projects to already existing ones, to see if they give enough improvement to justify their creation. The range of possibilities for future projects helps identify goals, requirements, and growth paths.

Part 2: Transport methods

The next part of the book considers transportation. Before you can do anything else at a given space location, you first have to get there. Historically the transportation component has been much more massive than the cargo. Solving transportation challenges became the primary focus of space projects. Although still important, we take a more balanced approach in this book. Improved transportation methods and use of local resources will shift the mass ratio towards the destination, so we give that part of space projects equal attention.

Many more transport methods have been proposed (about 75) than have actually been used to date (about 5), and one of them, **chemical rockets**, has been used by far the most. Part 2 lists the many known and speculative space propulsion methods by category. After listing the available concepts, the later sections will make some comparisons, and discuss how to select the best candidates for a given project.

Part of the reason chemical rockets have been used so much is "first mover advantage". They were the first type of space propulsion to get extensive development. They have the longest history, most optimization of design, and most familiarity, so they continue to be used. That does not mean they are objectively the best solution for all time, and all circumstances. Use of alternative methods, such as ion engines, is becoming more common in recent years. When starting a new project, a survey of all possible technical choices is worth doing, before narrowing the list down to the best candidates, so that no good option is overlooked. That is one reason the list in this book attempts to be comprehensive. Additionally, no single method is optimal for all locations and mission requirements.

Part 3: Engineering Methods

Part 3 considers what to do when you reach a location in space, and how to do it. It starts by listing overall design factors which affect multiple parts of a project. Having accounted for propulsion in Part 2, we then explore the other subsystems that make up a complete system item. A large program or project will consist of multiple system items with different purposes and functions. The remainder of Part 3 covers the engineering methods that are specific to these functions. This includes obtaining resources, turning those resources into useful materials and parts, assembly and construction, verification and test, operation and maintenance, and finally recycling and disposal. A particular method describes in general **how** a given task is done. Once completed, an engineering design then implements that method in a specific device.

The approach to date for most space projects has been to do all the design and construction on Earth, and then launch pre-built and pre-supplied hardware such as communication satellites or interplanetary probes as complete items. That is adequate for smaller projects, but as larger ones are contemplated this method becomes unreasonably difficult and expensive. The International Space Station is an example where it was too large to launch as one complete item, with supplies for it's full operating life. Instead it was assembled from smaller parts that each fit within a single rocket launch, and supplies are delivered periodically. This reduced the payload per launch to a manageable size, but all the parts and supplies are still being brought from Earth.

As larger and more distant projects are considered, the transport requirements continue to go up, and this starts to become very expensive. This is mainly due to the deep gravity well of the Earth, which requires a lot of energy to climb out of. At some point it becomes more economical for a project to extract supplies and energy, and eventually do production, locally. So we give a lot of attention to these tasks. Although doing them in space is new, there is a long history of engineering on Earth which can serve as a starting point.

Part 4: Complex Programs

Having looked in part 3 at the elements of a single system, Part 4 considers complex programs that are extended in time or involve multiple systems. A key point to understand is no single system works best in isolation. For example, a vehicle without a fuel supply can only travel once, with it's original fuel load. A fuel extraction plant can do nothing without raw materials to work on. Combining these, a vehicle can transport raw materials to the plant, and get fuel from the plant for later trips. Then the combined system can operate indefinitely as long as a surplus of fuel is produced from each trip. So to design a complete program, you need to look at all the component systems and how they interact.

In addition to the parts of a program interacting at a given point in time, it can also evolve over time. An analogy can be made to road systems on Earth. They started as footpaths, and were upgraded over time. The upgrades were not independent of the types of vehicles that used them, and the existing roads were used to deliver machinery and supplies to construct the upgraded roads. So it will be with sensibly designed large space projects - using one set of vehicles and facilities to help build the next level of improvements.

We give an extensive example of a complex program that both interacts with itself and grows over time. This example serves several purposes. One is to demonstrate by example how such a program is analyzed and defined. Where alternative options are presented, we explain how to choose among them. Another purpose is to serve as an actual proposed plan to implement in real life. In that context you have to go past pure design questions and consider cost and other external factors that apply to a real project. The final purpose is as a "class project" or "lab experiment" method of learning. Things like working in a team are best learned by doing, and using new knowledge in a project helps fix it in memory. Taking an unfinished part of our example to the next level of design detail can be used for this purpose. Experience working on a realistic design can be useful if someone intends to work in this field in the future.

Our proposed program is likely not the best one imaginable, but it is intended as a good starting point. As individuals and teams make suggestions and apply engineering effort to it we hope it will improve over time, both as an actual project proposal, and as a teaching tool.

Part 5: Design Studies

In this part we include or link to the details of design studies for elements of the program described in Part 4, and other studies done as examples. This allows the narrative in the earlier parts of the book to flow without interruption by a mass of detail, and also serves as examples of the types of studies and reports which engineers are frequently asked to work on. The full details showing the step by step assumptions and calculations are usually kept as part of a program's work history, and the results are distributed as reports for others to use.

References and Sources

We provide links to references and additional data in two main ways. References or sources that are specific to one method or idea are linked in the main body of the book. In-line links are boldface and capitalized. Additional references that cover multiple topics are included in this last section, along with appendices for data that does not fit in the main narrative. The first appendix, for example, lists fictional transport methods which do not have any scientific or engineering support. They are there for completeness, but as there are no practical prospects to use them, we put them in a separate section rather than the main body of the book. Additional Appendices include reference data.

Part 1: Science and Engineering Fundamentals

Introduction

The design and development of space systems is a part of engineering in general, which in turn relies on the knowledge base of mathematics and the sciences. Therefore we the first part of this book is a review of the fundamentals of these fields. We will pay the most attention to the parts that apply to space systems, but we by no means cover the whole of these fields of knowledge. A non-technical reader can get a general idea of the concepts and projects presented in this book, but it is mainly aimed at people who want a deeper understanding of, or to actually work with future space projects. To do that, a proper foundation in mathematics and the sciences is needed at a secondary education (high school) graduate/first year university science or engineering level. If you do not have such a background, there are open source textbooks available online, such as those from the **CK-12 Foundation**, as well as video lectures from the **Khan Academy**, and traditional books and classes.

Our discussion in this book is thus at an introductory engineering level. It is not a complete survey on any topic. In many cases there is simply too much detail to fit it all. In others the technical level is too advanced, and, in the case of some future methods, the ideas have not been fully developed yet. Other books, articles, and materials are linked throughout the book, especially in the **References** section, and also in our online **Library**. Readers are encouraged to delve deeper into any topics that interest them. The next few sections will give a more detailed summary of the background of mathematics and science, and how it relates to engineering and key design principles for space systems.

Mathematics

The importance of mathematics to science and engineering can be summarized in one sentence:

> **Our Universe appears to follow mathematical laws.**

By that we mean mathematical formulas and calculations produce results which match what we see when we look at the real world. This is a very powerful circumstance, because we can do the calculations **before** we look, even before something exists, and thus predict the future. Why mathematics works so well in describing reality is a philosophical question to which we don't have a good answer. This was pointed out by Eugene Wigner in 1960 in an article entitled ***The Unreasonable Effectiveness of Mathematics in the Natural Sciences*** (which is also discussed in a **Wikipedia article**). Regardless of why, it does work in practice. That allows us, among other things, to design systems that will work as intended.

The correspondence of mathematical predictions to the real world is not just a general one. In many cases it can be astoundingly exact. One of the earliest examples of prediction is the motion of the Sun, Moon, and planets in the sky. Even in ancient times people were able to predict where they would be in the future. Those predictions allowed knowing useful things, like when to plant crops, because of the linkage of the Earth's motion around the Sun to the seasons. Nowadays we can predict the motion of objects in the Solar System to fractional parts per million accuracy. An example of this accuracy was the 2012 landing of the Curiosity rover on Mars within 2 km of the intended location, after a trip of 566 million km. This could not have been done without predicting both the spacecraft trajectory and the future location of the landing point on a moving and rotating planet to 4 parts per billion. Further examples of using mathematics in design are all around you. Every tall building and bridge relies on the simple mathematical relationship that the strength be greater than the sum of all the loads. When you design such structures you calculate the strength, and calculate the loads, and then make sure the first is greater than the second. Proof that this method works is that tall buildings and bridges rarely fall down.

Like other engineering fields, space systems engineering relies on using such formulas and calculations. They are derived either from the sciences or practical experience and measurements within engineering. We present many of these formulas and calculations in this book. Therefore as a minimum you should understand the following mathematics topics (links are provided to introductory textbooks):

- **Algebra** - How to manipulate algebraic formulas and how to obtain a numeric answer given input values, the relationship of formulas and functions to graphs, and exponents and polynomials.
- **Geometry** - The types of geometric shapes and angles, and how the dimensions of two and three dimensional shapes (perimeter, area, and volume) are calculated.
- **Trigonometry** - Basic trigonometric functions and graphing them, vectors, and polar coordinates.
- **Probability and Statistics** - The ideas of averages, random error, distributions, and regressions.

More advanced topics, such as **Mathematical Analysis**, **Calculus** and beyond are helpful in understanding how the formulas are derived, or in solving the more complex problems in engineering. They are mostly not needed for an introductory level book such as this one.

Science

Science is a systematic enterprise that builds and organizes knowledge in the form of testable explanations and predictions about the universe. The predictions are often embodied in the results of mathematical formulas and calculations which relate to the real world. It is pursued partly on the basis that knowledge about the Universe satisfies human interest and curiosity. It also often turns out that the knowledge is useful in some practical way. We do not know in advance what knowledge will turn out to be useful, so scientists as a group study everything. Knowledge is a seamless whole, but from its history, and for the purposes of teaching and study, it is conventionally divided into branches according to the object of study. The ones most relevant to space systems include:

- **Physics** - This is the study of the forces and interactions of matter and energy, the results of those interactions, and the fundamental laws and components of which things are made.
- **Astronomy** - This is the study of objects and phenomena outside the Earth's atmosphere. This is the same location as all space systems operate in, so is highly relevant. Planetary Science in particular studies condensed objects orbiting stars.

- **Chemistry** - This is the study of matter at the atomic, molecular, and larger scales as far as how they react and their physical properties.

Other fields besides these three will also prove useful depending on the type of project. At least a basic understanding of these areas of science is needed to work with space systems, since engineering of those systems is derived from that knowledge. Beyond the branches of science, you should have an understanding of the **Scientific Method**, by which ideas are generated, experiments and observations are made to test those ideas, and thus they are validated or rejected. Peer review, statistics, and repeatability are among the methods used to ensure observations and conclusions are reliable. Absolute truth is never reached in science, merely increasing confidence in a given explanation, which is known as a **Theory**. Sufficiently well tested ideas join the body of knowledge considered settled, but they are always subject to revision and new ideas are constantly proposed.

Engineering

Mathematics and science are developed for their own sake and for their ability to predict the future. **Engineering** then applies accumulated knowledge, from the sciences and from experience, towards useful ends by designing, building, and operating systems to perform intended functions. When the systems are complex, a method called **Systems Engineering** is used across an entire project to organize and optimize the resulting design. This method can coordinate the work of thousands of people. Systems Engineering is described in more detail later in Part 1.

The total of accumulated engineering knowledge is too vast for any single person to know more than a small part of it. Therefore engineering in general is divided into major fields of specialization, each of which has it's own training path. It starts with a common basis in science and mathematics, then concentrates on particular areas of application, such as Mining, Chemical, Mechanical, and Electrical Engineering. Working engineers often further specialize their study and experience. They, or the organizations they work for, are called on as needed for each project. This is more efficient than keeping full time staff for every possible subject area. The specialists who are called on also have more experience in their area from having worked on many similar projects. Since the teams working on a project are not permanent, how you manage their interaction then becomes important. Project organization is also covered later in this part of the book.

Aerospace Engineering is the specialty field within which space systems fall. Space systems are projects which happen to operate in the space environment in the same way that ships and airplanes happen to operate in the water or through the air. Although the particular environment imposes differences in how things are designed, they all rely on the same base of knowledge in subjects like mechanics, materials science, and thermodynamics. So a complex project will use engineers from many of the specialty areas such as Mechanical, Chemical, and Electrical engineering, as well as Aerospace Engineers specialized in the methods and environments that apply to space. We will identify the other specialties later in Part 1 of this book, but will concentrate on the methods that apply to space. There are many existing books and articles about the other specialties for those who are interested.

This book is aimed at an introductory university engineering level reader. If you have no prior background in engineering or in space systems in particular, you may want to start with **Engineering - an Introduction** by the CK-12 Foundation. You can get additional background from some of the book-length and website references in the **References** section, the JPL **Basics of Space Flight**, Glenn Research Center **Beginner's Guide to Rockets**, and Mark Prado's **Permanent** and Robert Braeunig's **Rocket and Space Technology**

Design Principles

Through training and experience, engineers develop a sense of what will work or not, and how to optimize a design. Partly this is through broad principles that apply in their specialty. We note a few of the more important ones that apply to space systems here. These and others will appear throughout the book and we will try to highlight them:

- **Earth vs Space** - On Earth, transport involves friction of various kinds, and most things are moving slowly in relation to each other. Therefore energy and cost are proportional to distance, but not time. Space is a nearly frictionless medium, and things are moving at relatively high velocity with respect to each other. So difficulty and cost are more related to kinetic and potential energy, which governs the paths you follow. It also depends on the time you start, since your destination does not stay in the same relative location, rather than absolute distance.

- **Non-Linearity** - Many of the formulas and variables related to space systems have values raised to a power or an exponential. So the difficulty of a task does not have a one-to-one relation to the magnitude of the desired goal. This is called a **non-linear** system. Understanding the direction and amount of the non-linearity is important, as this can greatly help or hinder a given task. One of many examples is atmospheric pressure, which decreases exponentially with altitude, thus decreasing aerodynamic drag proportionally.

- **Uncertainty and Margins** - Although some values, like the orbit of a planet, are known quite accurately, no physical parameter is known with absolute accuracy. Anything built by humans will deviate by some amount from the ideal item embodied in the design drawings. The natural environment can fluctuate over time, and be uneven from measured averages. So all engineering designs need to account for the uncertainties in the physical data they are based on and production variations. One method to do this is to introduce **Design Margins** above the expected conditions that are larger than the uncertainties. How much margin to use is based on cost, experience, and the use to which the design is put. For example, a passenger airplane would generally have higher margins than a drone with no crew, even though both are aircraft.

Section 1.1 - Basic Sciences: Physics

Physics as a Subject

Physics is the study of how the Universe behaves in its component parts: matter, energy, forces, motion, space, time, and so on. Our understanding of physics is based on experiment and observation, and ideas developed to explain what we see. There are many ideas, but only one reality. So the **Scientific Method** uses experiment and observation to determine which ideas best match reality. Ideas are loosely graded in quality; as hypotheses, theories, principles, and laws. These terms are applied based on how firmly and widely they have been tested.

In the sciences, no idea is considered final or absolute truth. Rather, they are always subject to revision or replacement when confronted by new observations and experiments. Each new observation increases our degree of confidence for the ideas that it supports, and many ideas have been tested for so long, and in so many ways, that we can rely on them routinely, even in engineering projects whose failure would be catastrophic and expensive.

Some ideas in theoretical physics explore aspects of reality that are not easily observed. They require devices like larger and more expensive telescopes or particle accelerators, or nobody has figured out a way to test the idea yet. Theoretical ideas can be developed with nothing more than mathematics, a pencil, and pad of paper. However, most of physics is empirical - based things we can see, test, and use. For example, quantum mechanics is illustrated and tested by the double-slit experiment, and special and general relativity by observing bodies in motion through the solar system. These also have engineering consequences in some applications, such as integrated circuit design and the timing of satellite signals (such as GPS). Conversely, some observations, like rotation curves and gravitational lensing of galaxies, don't yet have a good explanation. We call what causes the observations **Dark Matter**, but we don't yet know what dark matter is. So physics and the other sciences are unfinished works in progress. Many parts are well understood and settled, but around the edges are parts still being worked on.

[Move to page 3 Astronomy: In some cases, it is possible to test something by looking for logical consequences which would not be true in competing scenarios. For example, measurements of the cosmological constant has helped narrow down the theoretical possibilities regarding the origins of the universe by excluding those which predict higher or lower energy levels.]

Scope of this section

In this section we discuss key physical principles that relate to space projects. These are only a subset of physics as a whole. We first look at them in ideal terms for the purpose of understanding them individually. However, realistic design work has to consider less than ideal conditions, such as friction or perturbing additional forces, and also must account for uncertainties in how well we know any measured property. This difference represents the difference between physical laws and practical engineering.

Physical principles are usually expressed mathematically as algebraic formulas and geometric relationships, with supporting explanations to provide meaning and context. When known numerical values with proper units are inserted into these formulas, you can solve for an unknown value you wish to know. The ability to calculate unknown values is enormously useful when designing or operating space projects. As noted in the introduction to Part 1, the reader should have an understanding of mathematics if they want to use these formulas themselves.

To Learn More:

It is not our purpose here to include an entire physics textbook, but rather a summary of the most important relationships that apply to space systems. For more detail on physics in general, you can refer to one of the following sources:

- Related books in the **Wikibooks Physics** subject heading,
- Physics articles listed in the Wikipedia **Outline of physics** and extensive **Index of articles**,
- Pages in the **HyperPhysics** website,
- Short YouTube videos from the **Khan Academy**,
- Open source textbooks like the ones at the **Light and Matter** or **Motion Mountain** websites, or
- A printed college textbook such as **University Physics**, though less expensive used textbooks may be a better choice.

Units and Coordinates

Units

In order to obtain the correct results from a formula, a consistent set of units and method of measuring physical quantities such as position in those units is necessary. For example, adding two feet to three meters to get five of something does not produce a meaningful result because the units are different. The **International System of Units** (PDF file), abbreviated from the French to **SI**, is the preferred system of units for engineering and scientific work. It is also known as the **Metric System** because the base unit of length is the "Metre" (or meter in English). For historical reasons some values in space systems design are reported in US customary units, but these should be converted to SI values. There are also units of convenience, such as **gravities** being a multiple of Earth's surface gravity. It is convenient to express acceleration effects on humans in this way, relative to the Earth normal value. It should always be recognized as a convenience, and converted to SI units when doing calculations by it's standard ratio of 1 gravity = 9.80665 m/s^2. Physical quantities include both the numerical value and the units, and units must be carried through properly when doing calculations.

The base SI units are not defined in terms of other units, but rather by a description of how to measure them from nature. In one case, the kilogram, the base is a physical artifact, but that may soon be replaced by measurements. The base units are currently the **Meter** for length, **Kilogram** for mass, **Second** for time, **Ampere** for electric current, **Kelvin** for thermodynamic temperature, **mole** for amount of substance, and **candela** for luminous intensity. Efforts are underway to define these base units in terms of constants of nature, but they are not complete. Derived SI units are the products of powers of the base units. For example, the unit of force, called the **Newton**, is 1 kilogram-meter times seconds^{-2}. Many of the derived units are named after famous physicists, but these named units are identical to the form expressed in base units. Multiples and sub-multiples of units are indicated by prefixes which indicate integer powers of ten ranging from -24 to +24. Among the more common are **kilo**, indicating 10^3 or 1000, and **milli-** indicating 10^{-3} or 0.001.

Position

In modern physics there is no absolute or preferred reference frame in the universe. Therefore position is measured relative to a starting point known as the **Origin**, which is given a value of zero. **Altitude** on Earth is measured relative to sea level in the direction opposite the local gravity direction. On bodies without an ocean to serve as a reference, an average ellipsoid based on the shape of the planet or satellite is defined as zero altitude. On gas giants, which do not have a visible solid or liquid surface, altitude is based on pressure. On near-spherical bodies, **Latitude** and **Longitude** are measured relative to the points where the surface meets the axis of rotation, which are called the **Poles**, and a point assigned values of zero in both coordinates called the **Zero Point**. Units of **Degrees**, which are 1/360th of a circle measure the location relative to the zero point.

Objects in space are generally moving in relation to each other in paths defined by gravitational forces. When the paths are purely the result of gravity they are called **Orbits** and are measured by six parameters called **Orbital Elements** from which you can calculate position at a given time. When absolute position is more useful, it can be measured in three dimensions relative to an origin, such as the center of the Sun, with axes defined relative to the stellar background. Alternately a radial distance and two angles relative to a reference plane can be used. Most often the reference plane is that which contains the Earth's orbit around the Sun, known as the **Ecliptic**. In a Universe of three physical dimensions, it takes three values to define a position uniquely, either distance in three axes, or a radius and two angles. These values are known as the object's **Coordinates**.

Motion

Displacement is the change in position. It has both amount and direction, such as "three kilometers North". **Velocity** is the rate of change in position per unit time. When stated without a direction and purely as an amount it has a single value in units of meters per second. Where x is position in the direction of motion, t is time, and the Greek letter delta (which looks like a triangle) indicates change in those values, then velocity v is given by the following formula:

$$\bar{v} = \frac{\Delta x}{\Delta t}$$

Acceleration is the rate of change in velocity, or the second derivative of position, with respect to time. Thus acceleration a is:

$$\bar{a} = \frac{\Delta v_x}{\Delta t} = \frac{d^2 x}{dt^2}$$

The horizontal bar over v and a in the respective formulas indicates velocity is directional. A value such as this when stated with both a magnitude and direction is called a **Vector**, while a value without a direction is called a **Scalar**. The direction can be given in terms of two angles, or the velocity can be expressed as components in the three (x, y, z) axes of a reference system, but either way a total of three values are required to state a velocity vector. **Vector algebra** is a method for doing calculations with vectors. It is somewhat different and more complex than simple algebra.

In accelerated motion, the velocity at any given instant is changing. We can define an **Instantaneous Velocity** at a particular time, and an **Average Velocity** over an interval. In a closed orbit the moving object returns to its starting point. Thus for a full orbit, the net change in position is zero, and as a vector the average velocity is also zero. If you measure the total length of the orbital path and divide by the time one orbit takes, you can obtain an average orbital speed as a positive scalar value. This illustrates how different vector and scalar values can be. Acceleration can also change with time. For example, the accelerating force due to gravity changes as the inverse square of distance. Thus a falling object will increase in acceleration as it gets closer. Under constant acceleration in a straight line we can determine change in position or distance d from:

$$d = \Delta x = 1/2 \times at^2$$

In circular motion, where v is the velocity and r is the radius, we can find the acceleration a from the following formula:

$$a = \frac{v^2}{r}$$

One use for this formula is finding a required velocity for a circular orbit from the acceleration of gravity (see under Forces below) and the radial distance from the center of the body. Note there is no centrifugal force. Gravity or a rotating structure provide an inwardly directed force to maintain circular motion of the object, but there is no outward one.

Forces

The 20th century theories of **General Relativity** and **Quantum Mechanics** are more accurate predictors of how objects behave in the realms of the fast and the small, but in many cases the simpler formulas of **Classical Mechanics** are sufficiently accurate to use. Examples where classical mechanics is not accurate enough include long term changes in the orbit of Mercury, which being the closest planet to the Sun, moves the fastest, and GPS navigation, which relies on extreme accuracy of the orbits of the satellites, and the effect of gravity on their signals, to determine user position. Isaac Newton formulated many of the basic ideas of classical mechanics in his **Principia**, published in 1687.These include his three laws of motion, conservation of momentum and angular momentum, and the law of universal gravitation.

Newton's Three Laws

These are laws in the mathematical sense, which Newton deduced from experiments performed by others before him. They involve two opposing concepts: **Forces** which tend to create motion, and **Mass** which tends to oppose it via the property of **Inertia**. The relationship of forces to the motion they create is known as **Dynamics**. Forces are vectors, having magnitude and direction, and multiple forces act as the vector sum of the component forces. This also means single forces can be decomposed into components, such as vertical and horizontal components relative to an axis system, or perpendicular (normal) and parallel components relative to a surface. Decomposition is done when it is useful in solving a problem. The three laws are:

First: Inertia - A body acted on by no net force moves with constant velocity (which may be zero) and zero acceleration:

$$\sum \vec{F}_i = 0 \Rightarrow a = 0$$

This is contrary to common earthly experience where friction acts to stop objects in motion.

Objects moving freely in the vacuum of space demonstrate this Law more clearly since it is a frictionless environment. An airplane in level flight has multiple forces acting on it (gravity, lift, thrust, and drag), but if the vector sum of all the forces is zero, it will continue moving at the same altitude and velocity in the same direction. From the fact that buildings typically are not accelerating we can deduce there is no net force acting on them. To put it another way, the sum of the forces is zero. Since gravity acts to pull the building down, there must be an equal force from the bedrock acting to hold it up. Applying this idea to every structural component of the building is a powerful way to determine the necessary design of those components - at each point where components connect, the forces must sum to zero, therefore you can calculate the forces which a particular component must withstand.

Second: Force - When a force does act on a body of mass m, the acceleration a is related to the magnitude of the force F by the formula

$$\vec{F} = m\vec{a}$$

The arrows above the symbols indicate they are vectors, meaning a quantity in a particular direction. Thus a force in a given direction produces an acceleration in the same direction. Manipulating this simple formula has very wide ranging use in space systems. Given any two of the values, we can find the third. Summing across time, we can find total velocity change.

Since mass has units of kilograms, and acceleration has units of meters per second squared, then by the above formula force has units of kilogram-meter per second squared. We call this unit a **Newton** (abbreviated "N") in the SI system of units, and is named after the scientist. The force which the Earth exerts on a falling apple is coincidentally, given stories about the scientist and falling apples, about 1 Newton. The force of gravity on an object is referred to as **Weight**. Most humans live where acceleration of gravity is within 0.2% of the same value, so we often confuse weight with mass. They are proportional, but they have different units. On another planet, the same object would have a different weight. Weight does not disappear when in orbit - aboard the Space Station the force of the Earth's gravity is only 10% less than on the ground. So-called "zero gravity" is more properly described as **Free Fall**. The astronauts inside the Station and the Station itself are both affected by the same acceleration of gravity. So the difference between them is zero, and the astronauts do not feel their bodies pressed against anything. On Earth what you feel is parts of your body pressed against the ground or furniture, and internally pressing against other parts of your body. This pressure is what you experience as "weight".

The product of mass times velocity is called **momentum**. It is given the symbol *p* since mass already uses the letter m. Force also equals the change of momentum with respect to time, since acceleration is the change in velocity with respect to time and we are just adding the multiplier of mass:

$$\vec{F} = \frac{d\vec{p}}{dt}.$$

Third: Reaction - Single forces do not act in isolation. At the most fundamental level the particles which carry the four forces of nature act on both the emitter and absorber of the particles. At the macroscopic level we live in, where forces are the combined action of many particles, we observe the dual action as for every force there is an equal and opposite reaction force. Where the subscripts indicate the force of object A on object B and object B on object A:

$$\vec{F}_{\mathrm{a}b} = -\vec{F}_{\mathrm{b}a}$$

Therefore a body can never move itself by applying forces only to itself, because the reaction force would cancel it out. You cannot lift yourself above the ground no matter how hard you try by applying forces to your own body. A pole vaulter, however, can raise their body a considerable distance by applying force to the ground. The reaction force of the ground through the pole then acts to raise their body. Of great interest for space systems is that a rocket engine applies a great deal of force to expel gases in one direction, and the gas applies a reaction force in the opposite direction, which moves the vehicle.

Multiplying both sides of the above formula by units of time, and subtracting the left side from the right side, we find that sum of momentum (mass times velocity) changes is always zero. This is known as the **Law of Conservation of Momentum**. It is referred to as a physical law because it has never been observed to be violated, and conservation in the physics sense means a value which does not change. It is found to be conserved both for linear and rotational motion. The latter is referred to as **Angular Momentum**. Thus the Earth would continue to rotate forever unless acted on by outside forces. Such forces do in fact act, mainly tidal forces from the Moon. So the Earth's rotation is slowing down measurably - the day is getting longer by 23 microseconds per year. But since angular momentum is conserved, slowing the Earth's rotation means the Moon increases its angular momentum. This increases the size of its orbit by a measurable amount (3.8 cm/year)

The Forces of Nature

There are only four fundamental forces that we know of, responsible for all motion in the Universe. These are the gravitational, electromagnetic, weak nuclear, and strong nuclear forces. These forces interact via gravitons, photons, W and Z bosons, and gluons respectively. For more detail see **Fundamental Interaction**. The latter two are short range forces which mostly occur within atomic nuclei, so the two that concern space projects the most are gravity and electromagnetism.

Gravity

Gravitons are the hypothetical particles which should carry the gravitational force. They are hypothetical because they have not yet been observed. Because gravitons never decay, their range is infinite, and the gravitational field of any object in the Universe affects every other object in the Universe. As a result, the total **gravitational field** surrounding an object remains the same at any distance. The area of a sphere surrounding an object is 4 × pi × the radius squared. So the gravitational field per unit area decreases with the square of the radius r. The rates of graviton production and absorption are both proportional to the mass of an object. Between any two objects the total gravitational force depends on the product of the two masses, since the first object emits a number depending on it's mass M, and then the second object absorbs some of them according to it's second mass m. The rate of graviton production and absorption is measured as a universal constant G which applies to every object in the Universe, as far as we know:

$$G = 6.67 \times 10^{-11} Nm^2/kg^2$$

Gravity always acts to attract two objects to each other, in other words reduce the distance, therefore the force is given a negative value. As a practical matter, since the field falls as the square of distance, objects sufficiently far away can be ignored to the extent you need to accurately calculate the total gravitational force on an object. The force acts on a line between each pair of objects and the total force is simply the sum of the individual forces accounting for the direction of each, and each is found by the formula

$$F = -\frac{GMm}{r^2}$$

Since force is also mass times acceleration, we can equate them and remove mass m from both sides of the equation, giving the acceleration due to gravity of an object with mass M as

$$\vec{a} = -\frac{GM}{r^2}$$

When restrained from accelerating, such as when you stand on the surface of the Earth, you experience the downward force as **weight**. Your mass does not change according to what object you are standing on, but the acceleration does, due to the object's different mass and radius. Therefore your weight will be different on other planets and satellites

When unrestrained from accelerating, also known as **free fall**, then the parts of your body, for instance if riding in vehicle, all will accelerate at the same rate. They do not have any acceleration relative to each other, sometimes called **Zero gravity**, but incorrectly since nowhere in the Universe is there truly no gravitational field. So in the case of unrestrained acceleration "free fall" is the more correct term, and "zero apparent gravity" will provide a better indication as to express how it appears to a human relative to their surroundings.

The local acceleration at the Earth's surface has a standard value of 9.80665 meters per second squared, but actually varies about 2% depending on location. The standard value is given the symbol g, and accelerations are sometimes stated as multiples of standard Earth gravity to give an impression of how humans would experience it, but for calculation purposes meters/second squared should be used to avoid unit errors. Similarly weight should not be confused with mass. Weight is in reference to the local gravity field, while mass is the moe correct unit to use at any location.

Electromagnetism

Photons, the particles which carry the electromagnetic force, behave similarly to gravitons in that they do not decay as they travel, and obey an inverse square field law. Where gravity is the result of mass, electromagnetic force is the result of electric charge. Unlike gravity, charge comes in two types which we call **positive** and **negative**. The names are arbitrary, positive charges are not larger or higher than negative ones, but they have the property that like charges repel each other, and unlike charges attract. The electromagnetic force is found by the formula

$$F = k_e \frac{q_1 q_2}{r^2}$$

Where F is the force, k(e) is fixed value called **Coulomb's constant**, q1 and q2 are the electric charges, and r is the distance between them. Note the form of this equation is similar to the one for gravitational force. When both of the charges are positive, or both negative, their product is positive, and so is the force. Positive forces act to increase the distance between charges. When the charges are unlike, one positive and one negative, the product is negative, and thus the force acts to decrease distance. Coulomb's constant, whee C is charge in units of 1 Coulomb = 1 mole of elementary chages is

$$k_e = 8.987 \times 10^9 \ \mathrm{N \cdot m^2/C^2}$$

Elementary charges are those on a single electron or proton, and are always observed as integer multiples of those charges, never fractions. Charges are additive by simple arithmetic, with negative charges canceling the fields of positive charges. Since unlike charges attract each other, they tend to annihilate if they are antiparticles or form neutral atoms if they are protons within atomic nuclei and electrons. So large quantities of matter tend to have low net charge. Since mass is always positive, large quantities of matter always have large amounts of gravity

Moving electric charges create a **magnetic field**, including the imputed spin of the charge from elementary particles. Materials with aligned atomic spins can have a static magnetic field. A steady flow of electric charges is called a **Current**, and also creates a field. Magnetic fields in turn affect the motion of electric charges creating a force F, where I is the current, l is the length of the wire, and B is the strength of the magnetic field:

$$\mathbf{F} = I\boldsymbol{\ell} \times \mathbf{B}$$

The bold face symbols indicate these are **vector** values, having directions. The force is perpendicular to both the direction of the wire/current and the magnetic field. Natural magnetic fields, such as the Earth's, are assumed to be caused by electric currents within the metal core.

Continued on page 2 →

Section 1.1 - Basic Sciences: Physics

(page 2)

← Back to Page 1

Energy

Work in the physics sense is a force applied through a distance, or in equation terms W = Fd. It is a scalar (numerical) value found by multiplying two vectors (directional values), the direction of the applied force, and the direction of motion. Since those directions do not have to be the same, the vector product can vary as the cosine of the angle between them, and therefore be zero or negative. For example, if you apply a lifting force to a table, but not enough to raise it off the floor, you do no work in the physics sense, even though your muscles will tell you they are working in the biological sense. If you manage to lift the table, the direction of motion (up) is opposite the direction of gravity (down), and therefore the work done on the table is negative. As odd as that sounds in conversation, the mathematics works out when solving physics problems.

Energy is then defined as the ability to do work. It comes in many forms which can be converted either by natural actions or human devices.As far as we have reliably observed, total energy always remains the same, a principle known as **Conservation of Energy**. An exception to this might be **Dark Energy**, a hypothesis to explain the apparent acceleration in expansion of the Universe. We do not yet know what Dark Energy is, though. It is a label we apply to explain certain observations, and we definitely have no way to apply such energy. For practical engineering purposes we ignore it and treat conservation of energy as a firm principle.

Energy is measured in SI units by **Joules**, named after a 19th century physicist who helped discover the relationships of energy, work, and heat. Since energy comes in different forms, the Joule has several equivalent definitions.Leaving out the numerical values and only looking at units it can be expressed as:

$$\mathrm{J} = \frac{\mathrm{kg} \cdot \mathrm{m}^2}{\mathrm{s}^2} = \mathrm{N} \cdot \mathrm{m} = \mathrm{Pa} \cdot \mathrm{m}^3 = \mathrm{W} \cdot \mathrm{s}$$

The first equivalence is to base SI units of kilograms, meters, and seconds. Where N is Newtons, Pa is Pascals, the unit of pressure, and W is Watts, the unit of power, the following expressions are in terms of force × distance, pressure × volume, and power × time. Note that W as work and W as Watts mean different things, and the latter is distinguished by having a quantity attached (ie 100 W meaning 100 Watts). Unfortunately there are more physics concepts than letters of the alphabet, which can be confusing at times. When a formula could lead to such confusion, write out the unit in full rather than abbreviate, or define the symbol in words as we usually do around a formula. Which of the above expressions for energy are appropriate to use depends on the types, and which conversions of it, are involved in a particular situation. One of the forms which energy takes is as matter. Where E is energy, m is mass, and c is the speed of light they are related by the famous equation

$$E = mc^2$$

Since the speed of light is a large number, by definition exactly 299,792,458 meters per second, and that number is squared in the formula, the energy contained in a given amount of mass is enormous. The conversion of less than 1% of mass in nuclear reactions produces sufficient energy to power stars, atomic bombs, and nuclear power reactors. Objects which are moving in gravity fields are very common in space projects. We find it useful to define the following two energy quantities based on their motion and position:

Kinetic Energy

Kinetic Energy is the energy an object possesses through its motion. It can also be described as the amount of work required to get a body of mass to move. In mathematical form Work = change in Kinetic Energy, or

$$W = \Delta KE = K_2 - K_1$$

This is in addition to the bound energy of matter (mc^2). Referring to Newton's first law of motion, an object will retain its kinetic energy unless acted upon by another force such as friction or gravity. For example, objects free of a gravitational field and in a vacuum will retain their kinetic energy, direction, and velocity. Kinetic energy KE is a function of mass m and velocity v according to the formula

$$KE = \frac{1}{2}mv^2$$

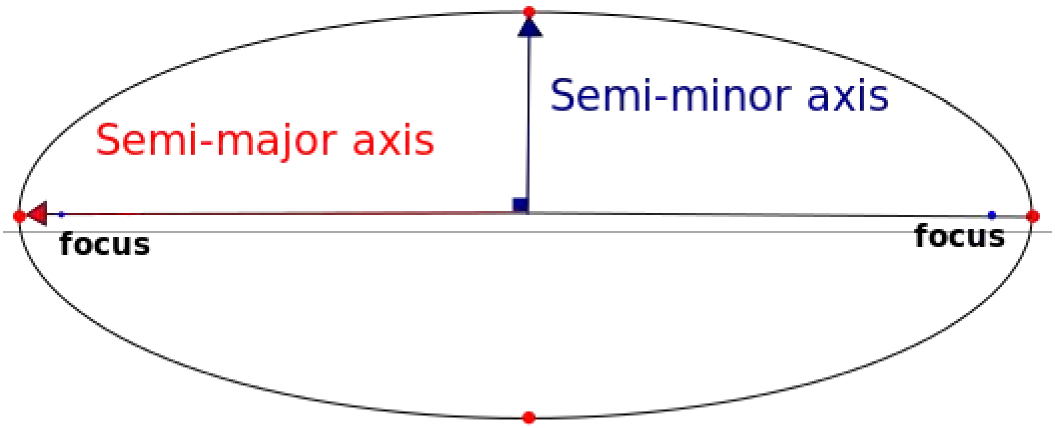
A **Reference Frame** is a non-accelerating environment. Velocity is measured relative to such a reference frame. Therefore a space station in orbit, and an astronaut inside it, may both have a large velocity relative to the center of the Earth. They then both have a large kinetic energy in an Earth-centered reference frame, in fact sufficient to raise their temperature hotter than the Sun, to 7000 K. Relative to each other, however, their velocity is near zero, so in a reference frame moving with them they have near zero kinetic energy. Since the formula above takes the square of velocity, kinetic energy is always positive, even if the velocity is negative in a given reference frame.

Potential Energy

Potential Energy is the difference between the energy of an object in a given position and its energy in a reference position. When work is done against a conservative field, such as gravity, then the energy of that work is converted to potential energy. Setting the reference position as infinitely far away, the gravitational potential is always negative since you must do positive work to lift an object to infinity. Since gravity varies as the inverse square of distance, the sum of the work going to infinity varies as the inverse of distance. Where G is the gravitational constant, m(1) is the mass being moved, and M(2) is the large mass producing gravity, then the potential U is:

$$U = -G\frac{m_1 M_2}{r}$$

If no forces besides gravity are acting on an object then the sum of kinetic and potential energy is constant. Colliding with a planet or other object involves other forces than gravity. So an object in an elliptical orbit, which does not collide, and is free to repeat its motion, constantly exchanges potential and kinetic energy as the distance r from the center of the body it is orbiting changes. It has more kinetic and less potential energy at the lowest point and so moves faster. The velocity v at any point in the orbit can be found from:



$$v = \sqrt{GM\left(\frac{2}{r} - \frac{1}{a}\right)}$$

where:

- r is the distance between the orbiting object and the body it orbits. The body is located at one focus of the ellipse.
- a is half the long axis of the orbit shape, or **Semi-major Axis** (see figure at right)

Mechanics

Mechanics is the description of the motion of an object under the influence of forces. For space projects this is usually the **Thrust** forces generated by a propulsion system, and the influence of **Gravity**. The motion in a vacuum among massive objects like planets and the Sun is called **Orbital Mechanics**, which we will cover in section 1.2. When operating within an atmosphere, an additional force is generated when moving. This force is decomposed into a perpendicular component called **Lift**, and a parallel component called **Drag**. When two objects are in contact they additionally generate a force which is decomposed into a parallel, or **Friction** force, and a perpendicular, or **Normal Force**. The combination of all forces, including less common ones not listed in this paragraph, produce a vector sum total force on the object and thus an acceleration in some direction.

Friction

Friction forces matter for space systems for things like rovers moving on the surface of a body and for internal parts of motors and pumps. Since they always oppose motion, they require energy to overcome, and thus a source of energy in the system. Friction at a microscopic level is caused by interactions between electrons of the two objects, thus is an electromagnetic effect. At a slightly larger level they cause temporary bonding and physical obstruction due to surface roughness. At human scale these microscopic interactions can be summed up as average values which are proportional to the normal forces and depend on the types of surfaces in contact. The multiplier to the normal forces is called the **Coefficient of Friction**, with symbol μ. The friction force, f, is then related to the normal force, F_n by:

$$f = \mu_s F_n$$

Coefficients of friction are determined experimentally, and depend on whether the objects are moving (kinetic friction) or not (static friction). Static friction is typically higher because the objects have time to form atomic bonds and settle into the bumps of surface roughness. The coefficients also depend on type of material and whether any gas or liquid is trapped between them. The ability of skaters to move easily on ice comes from a microscopic layer of water which forms due to pressure of the blades. Vertical lift from contact only requires breaking atomic bonds, and not interlocking of surface roughness. Thus wheels and ball bearings, which vertically separate the contact surfaces, have lower **Rolling Friction** than sliding contact.

Normal Forces

A **Normal Force** acts perpendicular to a surface, and has several sources. (Normal here refers to the perpendicular direction in geometry, not common or average.) They includes gravity, magnetic or electrostatic attraction, and gas or liquid pressure. In reality, friction and normal forces are components of the total contact force. Since motion is prevented in the perpendicular direction by the existence of a solid surface, it is easier to calculate the effects by looking at the components separately. When perpendicular motion is **not** prevented, which happens with liquids and gases, it becomes more complex. The field of **Fluid Dynamics** is the study of these more complex motions, both against solid surfaces and internally within fluids.

Thrust

Thrust is the force generated *by* a vehicle by expelling reaction mass or by interacting with the environment. When something external acts *on* the vehicle it is referred to as an **accelerating force**, and often specifically named. The magnitude of the thrust due to expelled mass is given by

$$\mathbf{T} = \frac{dm}{dt}\mathbf{v}$$

where **T** is the thrust generated (force); $\frac{dm}{dt}$ is the rate of change of mass with respect to time (mass flow rate of exhaust); and **v** is the speed of the exhaust measured relative to the vehicle.

Drag

Drag is a force component generated by interaction with a fluid medium, such as the Earth's atmosphere. It is parallel to the incoming flow direction, and given by the formula:

$$F_D \,=\, \tfrac{1}{2}\,\rho\,v^2\,C_D\,A$$

where F_D is the drag force, ρ is the mass density of the fluid, v is the velocity of the object relative to the fluid, A is the reference area, which is the projected area occupied by the vehicle in a plane perpendicular to the motion, and C_D is the **Drag Coefficient** — a dimensionless number. While most of the terms in the above formula are simple to determine, drag coefficient varies in a complex way based on object shape, velocity, and other parameters. This is caused by complex flow conditions such as turbulence, shock waves, heating, and even decomposition at higher velocities.

When a surface moves relative to a fluid, the layer closest to the surface is affected most by the molecules colliding with the surface. They are deflected by the angle of the surface, roughness in the surface, or atomic forces between their respective atoms. This causes that fluid layer to tend to move along with the surface. The surface layer in turn affects farther layers by collisions of the molecules. At lower velocities, this sets up a smoothly varying **Boundary Layer** near the surface. At higher velocities the deflection is violent enough to create flow vortexes, where the fluid develops circular motions perpendicular to the direction of motion, known as **Turbulence**. Since it takes more energy to create the vortexes, the forces on the surface are higher, increasing friction or drag. These effects happen both externally to a vehicle moving in an atmosphere, and internally to a gas or liquid flowing within an engine.

The ratio of inertial forces, such as the sideways deflection, to the viscous forces caused by shearing (varying speed) in the boundary layer is called the **Reynolds Number**. The transition from smooth, or **Laminar Flow**, to turbulent flow, and the size of the vortexes, and thus the drag, is found experimentally to depend on Reynolds Number. It is a **Dimensionless Number**, meaning all the units in the formula cancel out when using consistent units, leaving a pure number. The Reynolds Number, Re, is found by:

$$\mathrm{Re} = \frac{\rho \mathrm{v} L}{\mu} = \frac{\mathrm{v} L}{\nu}$$

where: v is the mean velocity of the object relative to the fluid (m/s), L is a characteristic linear dimension of the surface (m), μ is the **Dynamic Viscosity** of the fluid (Pa·s or N·s/m² or kg/(m·s)), ν is the **Kinematic Viscosity** ($\nu = \mu/\rho$) (m²/s), and ρ is the density of the fluid (kg/m³). The characteristic dimension is defined by convention for various types of shapes, such as the diameter for a sphere. Since the motions within a turbulent fluid are too complex to reduce to a simple formula, for early design purposes drag coefficients are usually found from tables and graphs based on Reynolds number, which in turn were developed from experiment or historical data.

In more detailed or important design projects fluid forces like lift and drag are measured for the proposed design in a wind tunnel or other experiment, or calculated by detailed numerical simulations, a topic known as **Computational Fluid Dynamics**, or CFD. In CFD simulations the flow is broken up into sufficiently small volumes that the flow in each volume obeys relatively simple formulas, and thus the total flow in the simulation can be determined reasonably accurately. Historically this needed the largest available computers, and therefore physical testing often proved easier. With the vast increase in computer speed in recent decades this method has become much more practical.

Lift

Lift is the other force component generated by interaction with a surrounding medium. It is perpendicular to the incoming flow direction, and given by the formula

$$L = \frac{1}{2} \rho v^2 A C_L$$

where L is lift force, ρ is density of the medium, v is the velocity relative to the medium, A is planform area of the shape, and C_L is the lift coefficient at the desired **Angle of Attack** (angle between the reference plane of the shape and the velocity direction). Lift Coefficient depends on **Mach Number**, and **Reynolds Number**. Like drag, lift coefficient is a result of complex fluid flow, and depending on the design need is found by looking at a table or graph, physical experiment, or CFD simulation.

Thermodynamics

Thermodynamics is a branch of physics concerned with heat and temperature and their relation to energy and work. For space projects it becomes important for things like the operation of rocket engines, aerodynamic heating during reentry, and thermal radiators to dissipate excess heat. Thermodynamic variables and mathematical laws are simplifications applied to bulk amounts of material. Their underlying cause is the microscopic behavior of a very large number of smaller particles like molecules. Thermodynamics is often related to chemistry, because chemical reactions can give off or or absorb heat. A prime example is combustion in chemical rocket engines, where the propellants (fuel and oxidizer) react to produce very high temperature gas. Thermodynamics is a complex subject, so we cannot delve into all aspects of it here. We refer you to full textbooks like **Thermodynamics and Chemistry 2nd edition** by Howard DeVoe.

To give you a flavor for the topic, and because chemical rockets are very important in current space projects, we will describe their combustion cycle here, but it is not a complete description and many specialized terms are used:

The thermodynamic cycle for a liquid rocket booster is a modified Brayton (jet) cycle. A one-dimensional analysis may be performed by assuming the following ideal steps.

1. Fuel is injected into a combustion chamber isentropically either through use of pressurized fuel tanks or by a high-pressure pump, increasing the pressure to p_c and increasing the enthalpy
2. Heat is added to the fuel by means of combustion In an ideal situation, it is assumed that the pressure remains constant during this step, but that the temperature rises Both enthalpy and entropy increase during this step.
3. The combusted fuel expands isentropically to the exit pressure p_e, as it goes through the nozzle into the surroundings, which is at pressure p_0. Ideally p_e should equal p_0. During this process, the enthalpy decreases from h_c to h_e.

The thrust produced by a rocket is given by

$$T = \dot{m}_p v_e + A_{exit} * (p_e - p_0)$$

where $\dot{m}_p$ and v_e are the mass flow rate and exit velocity of the propellant, A_{exit} is the exit area of the nozzle and p_e and p_0 are the pressure at the exit point of the nozzle and the atmospheric pressure. The enthalpy represents the internal energy available for work or the potential energy. Thus, the energy change per unit time as the propellant moves from the combustion chamber to the nozzle exit is

$$\dot{m}_p(h_c - h_e) = \frac{1}{2}\dot{m}_p v_e^2$$

Solving for the propellant velocity yields

$$v_e = \sqrt{2(h_c - h_e)}$$

Let us assume that the combustion mixture of the propellants is an ideal gas. The internal energy per unit mass of an ideal gas is given by

$$h = \hat{c}_V T$$

producing an equation for the propellant velocity of

$$v_e = \sqrt{2\hat{c}_V(T_c - T_e)} = \sqrt{2\hat{c}_V T_c\left(1 - \frac{T_e}{T_c}\right)}$$

When an ideal gas expands isentropically, a change of temperature and pressure such that the following two relations hold

$$\frac{p_1}{p} = \left(1 + \frac{\gamma - 1}{2}M^2\right)^{\gamma/(\gamma-1)} \text{; and}$$

$$\frac{T_1}{T} = 1 + \frac{\gamma - 1}{2}M^2;$$

where M represents the Mach number at the location having static pressure p and temperature T. Using these two equations, we can relate the temperature and pressure ratios as

$$\frac{T_1}{T} = \left(\frac{p_1}{p}\right)^{(\gamma-1)/\gamma}$$

Thus, we can rewrite the equation for the propellant velocity as

$$v_e = \sqrt{2\hat{c}_V(T_c - T_e)} = \sqrt{2\hat{c}_V T_c \left[1 - \left(\frac{p_e}{p_c}\right)^{(\gamma-1)/\gamma}\right]}$$

The final analysis step in the one-dimensional analysis is the effects of the nozzle. The previous equation demonstrates that making the ratio p_e/p_c as small as possible maximizes the propellant speed, which in turn maximizes the thrust. The nozzle is designed to match the exit pressure as close as possible to the pressure of the atmosphere or the vacuum of space.

Continued on Page 3 →

Section 1.1 - Basic Sciences: Astronomy, Planetary Science, Chemistry (page 3)

← Back to Page 2

Astronomy

Astronomy is the study of all objects and phenomena beyond the Earth's atmosphere, and **Planetary Science** is specifically the study of condensed objects orbiting stars. Since this is where space systems function, a basic understanding of these fields is highly relevant to working on such projects. We provide only a very short introduction here. For more background than will fit in this book, see Nick Strobel's **Astronomy Notes** and the Wikibook **General_Astronomy**. One of the key ideas to emerge from these studies is the **Uniformity of the Universe**. As far as we can tell, the natural laws and processes that operate here and now have always operated in the past, are the same everywhere in the Universe, and we expect them to continue to be so in the future. Having learned what these general principles are, we can then apply them to specific examples as needed.

Origin and Evolution of the Universe

The **Universe** is the totality of existence. The origin and history of the Universe as a whole is of great interest to many people for its own sake, but only selected features are relevant for space systems design. This includes that **Baryonic** matter (the ordinary kind of matter that we and the Earth are made of) started out as about 76% Hydrogen, 24% Helium, and almost nothing else. Gravity caused the nearly uniform early Universe to develop denser regions with emptier regions in between. The denser regions coalesced into many **Galaxies**, of which the **Milky Way** galaxy is the one which our Sun and planets are part of.

Galaxies in turn form denser condensations where nuclear reactions occur, which we call stars. The reactions convert lighter elements into heavier ones, increasing the proportion by mass of Helium to about 27% and heavier elements to about 2%. Stellar nuclear reactions release a great deal of energy, but this source is finite. So the Sun and other stars will eventually run out of fuel, and the composition of the Universe will reach a stable condition. Several lines of evidence indicate the current age of the Universe is about 13.6 billion years, and the era of stars will last about 100 trillion years. If the expansion of the Universe continues to accelerate, then most of the Universe will be rendered undetectable long before all the nearby stars die out. Regardless of the eventual destiny of the Universe, on human time scales it will last a long while relatively unchanged.

The Milky Way Galaxy

The oldest stars in our home galaxy are about 13 billion years old, which indicates the galaxy started forming, at least to the point that stars condensed, shortly after the Universe as a whole formed. The baryonic mass is embedded in a five times more massive amount of material which only reveals itself by its gravity. Since this material does not form stars, it is dark, and so we call it **Dark Matter**. It is poorly understood at present, and our main concern for space projects is how it affects the motions of baryonic objects such as stars, planets, and molecular clouds. Evidence from the composition and motions of parts of the Milky Way indicate it formed by infall of gas clouds and smaller galaxies, which continues to the present. The shape seems to have evolved starting with the halo and central bulge, followed by growth of the disk shaped region. The baryonic mass of our galaxy is estimated at 200-300 billion times the Sun's.

The Sun is in the disk region, about 27,000 light years from the center of the Milky Way, orbiting at about 220 km/s and thus taking about 225-250 million years to complete an orbit. We do not know the exact shape of the Sun's orbit but it is suspected to be elliptical. Random motions of nearby stars are on the order of 50 km/s. Over the age of the Sun these random motions amount to 450,000 light years, which is much more than the circumference of the Sun's orbit. This indicates the current nearby stars are not the ones the Sun was born near. In fact, the current stars within 100 light years will be replaced by an entirely different set in 1 million years.

Formation of the Sun

We observe new stars forming in denser regions of our galaxy known as **Molecular clouds**, and we assume our Sun and the rest of the Solar System formed in such a cloud, which has since dispersed. Based on radioactive dating we estimate our Solar System to be 4.6 billion years old, which is about 1/3 the estimated age of the Universe. The presence of 1.5% heavier elements in the Sun confirms that it formed from recycled matter that had previously been enriched by older generation stars. Loss of heat from radiation allowed gravity to collect part of the original molecular cloud into a distinct object called the **Solar Nebula**. The core of the nebula continued to contract, and the increased pressure caused by self-gravity heated that core to create a **Proto-Sun**. Once the core of the proto-Sun reached a temperature of 12 million kelvins, hydrogen fusion could begin, and the Sun proper was born. This collapse until ignition took around 30 million years.

Nuclear reactions in the core of the Sun have converted Hydrogen to Helium, increasing the concentration there to about 60%. Since Helium is heavier, the core has gotten denser and hotter, thus increasing the reaction rate of the remaining Hydrogen and the total energy output of the Sun by about 40%. The current output is 3.846×10^{26} Watts. This will continue to increase by about 1% per hundred million years.

Planetary Science

Solar System Evolution

Planetary Formation

Whatever internal motions the Solar Nebula had, there was no way to dispose of angular momentum (net rotation). Therefore a small part of the nebula remained orbiting the proto-Sun rather than falling in, and keeping most of the angular momentum. This region is estimated to have been 50 AU in radius, and disk shaped as the net result of rotation. The increasing temperature of the proto-Sun created a temperature gradient based on distance from the center. The innermost part was too warm for icy material to remain solid, while the outer parts were cold enough for water, ammonia, and other ices. No part was cold enough for Hydrogen and Helium to condense. Particles condensed out of the nebula as the flat shape radiated heat to space and the optical thickness radially kept the outer parts from being heated by the Sun. Small particles could grow first by sticking to each other, then later by gravitational attraction. The mix of objects which formed this way are called **Planetesimals**. We now have fairly good evidence of this process by observing disks around young stars.

Gravitational attraction is a runaway process. As an object gets larger, it can attract objects from a larger distance, thus increasing its growth rate. Larger objects also have a potential energy well, so approaching objects will accelerate to impact. The impact energy eventually becomes large enough to melt the object. In addition, there were more radioactive elements in the early Solar System than there are now, and decay of those elements contributed to the heating of the growing bodies. The largest objects were able to affect the orbits of the smaller ones, causing them to either impact or get scattered away. This tended to clear out a region around each large object. The very largest objects had a sufficiently deep gravity well that they could collect gaseous Hydrogen and Helium, forming the gas giants. Some of the scattered objects, and planetesimals which formed at the outer edges of the nebula, have survived relatively unchanged beyond Neptune. The material from the inner Solar System which could not condense there tended to be blown outwards, and the point where they could solidify is near Jupiter's orbit, which may account for the large mass of Jupiter and Saturn compared to the rest of the planets. The entire accretion and clearing process took about 100 million years, with the final growth of the planets perhaps taking 10 million years.

Planetary Evolution

The bodies which were large enough to melt had their denser compounds sink to the center under their own gravity. Iron and related metals are the heaviest common materials, so they ended up in the cores. Going outwards, the layers include rocky minerals of different densities (a mantle and crust), then ices and an atmosphere if the body formed with these components. This layered structure is what we find today in the planets and larger satellites, along with the composition change with distance from the Sun. These trends are not strict rules because random collisions and gravitational scattering have changed the location and make up of the objects since they formed, and smaller bodies have lost original atmosphere and in some cases ices. After the original formation era of about 100 million years, the larger planets continued to interact with each other chaotically until they settled into a relatively stable arrangement about 3.8 billion years ago. The planetary shifts affected the smaller bodies, who continued to be scattered or impact. The evidence of this is still visible in craters and the locations of scattered objects.

Bodies in the Solar System continue to interact gravitationally, and impacts and scattering continue, but at a lower rate. Bodies like the Earth have active processes that tend to erase craters. Smaller ones lack an atmosphere or crustal motion, and preserve them through the life of the Solar System. The resulting current distribution of matter, and the very large energy output of the Sun, are the main resources to work with for space projects.

The Earth-Moon System

Earth Science is the study of the Earth and its component parts. The study of the Earth predates detailed study of other planets because this is where humans started, and it continues to be the best studied planet. In the context of modern astronomy and planetary science, the Earth is now studied as one planet among many. In the context of human history it still has a special place because we evolved here, and until now, none of us has left the Earth-Moon system. Nearly all of the design, materials, equipment, and operations for space projects to date has actually occurred on Earth. This will continue to be true for at least the near future. Therefore some understanding of the Earth is still needed to carry out space projects. For a longer introduction to the field see:

- CK12 Foundation - **Earth Science Concepts**

The Earth formed in the same way as the rest of the large bodies in the Solar System, mostly by collisions. Debris from a very large collision late in the process formed the Moon, explaining the difference in its composition relative to the Earth. Impacts and radioactive decay released enough energy to melt the entire planet, and the high temperature likely led to loss of some of the more volatile ices and gases. Continued radioactive decay, supplemented by other energy releases, has kept the Earth's interior hot. It has an inner metallic **Core** solidified by pressure, even though it is about the same temperature as the

Sun's surface (6000 K). Outside of this is a liquid metallic outer core, and then a rocky layer called the **Mantle**. The deeper parts of the Mantle have temperatures of 2000K or more. Although pressure keeps them solid at these high temperatures, the rock is able to flow slowly over time in a type of thermal circulation taking on the order of 100 million years. The least dense and coolest rocks form a roughly 120 km thick solid layer called the **Lithosphere**, which has a relatively high temperature gradient relative to the rest of the planet. By composition the lower part of the Lithosphere is part of the Mantle, and the outer part is less dense rock called the **Crust**.

The internal motions of the Mantle and heat traveling outwards can cause local temperature to get higher than the melting point dictated by local pressure. The molten rock is called **Magma**, and its composition can vary because different minerals have different melting curves. Movement of magma and bulk Mantle circulation drive about 20 pieces of the Crust, called **Plates** to slowly move across the Earth's surface and change shape. These movements and **Weathering**, the mechanical and chemical changes from the surface environment, explain the geography and geology we find today. These dynamic processes combine to erode most of the early Earth's history. The current surface averages ten percent or less of the age of the planet as a whole.

The Moon

Early in the Earth's history, a Mars-sized body called **Theia** is presumed to have hit the proto-Earth. Some of the debris from that impact collected by gravity to form the Moon. It formed much closer to the Earth than it is now, but tides act to slow the Earth's rotation and increase the Moon's orbit, a process that continues. The Moon is smaller than the Earth, and so lost its internal heat faster, and is now mostly solid. It is too small to retain an atmosphere because of the low escape velocity. So the Moon retains evidence of its early history in the form of large impact basins and craters of all sizes. The larger basins filled with magma to create relatively flat and darker areas mistakenly named **Mare** (Latin for "Sea") because, before telescopes, they were thought to have water. The greater tides from the time the Moon was closer to Earth slowed the Moon's rotation so the same side now always faces us, with a little wobble. It also caused most of the Mare to form on the near side.

Chemistry

The science of **Chemistry** has historically been considered a separate subject from physics. In a more general sense it can be considered a subset of low energy physics where arrangements of atoms via atomic bonding is important. We humans happen to require living conditions where atomic bonding is important, so we give that energy regime more attention. In reality, something like 99% of the matter in the Universe is in the plasma state, where electrons are no longer bound to atoms, and inter-atom bonding is rare.

The importance of chemical reactions to space projects until now was primarily in providing high temperature gases to propel rockets. In future projects chemical reactions for life support systems, and extraction and preparation of raw materials in space, will become much more important. So at least a basic understanding of chemical principles is useful for space systems design. We refer the reader to the following sources for more detail:

- CK-12 Foundation **Chemistry** pre-university level textbook mentioned at the start of Part 1
- Wikibooks **Introductory Chemistry**
- **Chem1 Virtual Textbook** by Stephen Lower

Section 1.2 - Orbital Mechanics

Introduction

Astrodynamics or **Orbital Mechanics** is mainly concerned with motions under gravity, either purely as a single force, or in combination with forces like thrust, drag, lift, light pressure, and others. As a topic it has a long history, with the motions of the planets, Moon, and Sun studied since ancient times, and with a scientific base starting about 500 years ago. With the advent of human-built spacecraft it has shifted from merely observing the motions of natural bodies, to planning and executing missions. The relevance to Space Systems engineering is, of course, the need to travel to a desired destination or orbit.

We will present some of the key ideas here. A more detailed and advanced introduction can be found in the Wikibook **Astrodynamics** and a set of Wikipedia articles at **Astrodynamics - A Compendium**. A good introductory printed textbook is **Fundamentals of Astrodynamics**, and there is an **MIT Astrodynamics** open course with downloadable materials.

Orbits

Gravity has no limiting distance, and therefore every object in the Universe is affected by the gravity of every other object in the Universe. For practical purposes, the force of gravity from most of the Universe cancels out, since there is about the same amount of material in all directions. What remains are more massive and nearby objects, whose gravity is large enough to matter for a given calculation. Which objects matter depends on how accurate your result needs to be. Typically detailed calculations are done with a computer simulation, since objects in general follow orbits, and move in relation to each other. Therefore the strength and direction of their gravity varies.

An **Orbit** is the path that an object will follow when only affected by gravity. Orbits around uniform single bodies are **Conic Sections**, which are shapes generated by slicing a cone. In order of eccentricity these are circle, ellipse, parabola, and hyperbola. Circular and elliptical orbits are bound to the body being orbited and will repeat. Parabolic and hyperbolic orbits are not bound to the body, although influenced by its gravity. They will not repeat. Simple orbit calculations only consider the nearest massive body. This is suitable when that body's attraction is much greater than other bodies, and for short time periods. More detailed and accurate calculations have to consider non-uniformity of the main body, and other bodies with enough gravity to influence the accuracy of the result.

First let's consider the ideal case of a single uniform massive object being orbited. Circular orbits have a constant velocity and distance from the center of mass of the body. This also means they have a constant **Orbital Period**, the time to complete one revolution around the body and return to the starting point. The circular orbit velocity, v_o, for any body can be found from:

$$v_o = \sqrt{(GM/r)}$$

Where G is the Gravitational constant (6.67 x 10^{-11} Nm^2/s^2), M is the Mass of the body orbited (in kg), and r is the radius to the center of the body orbited (in meters). G is a universal constant, and the mass of the Earth is essentially constant (neglecting falling meteors, atmosphere leakage, and things we launch away from Earth), so often the product GM = K = 3.986 x 10^14 m^3/s^2 is used.

The orbital period, P of a small body orbiting a central body in a circular or elliptic orbit is

$$T = 2\pi\sqrt{a^3/K}$$

Escape velocity, the velocity required to escape from a body's gravity to infinity, or v_e is found by

$$v_e = \sqrt{(2GM/r)}$$

Since this formula is the same as that for circular orbit, except by a factor of 2 in the square root term, escape velocity is the square root of 2 (1.414+) times circular orbit velocity. Elliptical orbits will have a velocity at the nearest point to the body, or **periapsis**, in between that of circular and escape.

Orbital Elements

Orbital Elements are the parameters required to fully describe the location and orientation of an orbit, the shape of the orbit, and the position of an orbiting object at a given time. They are described relative to the major body the object orbits. These elements change over time under the gravitational influence of other objects, which is called

Perturbation, non-ideal shape and mass distribution of the major body, the effects of relativity, and outside forces like drag and light pressure. The more important elements include:

Axes - Periodic orbits are generally ellipses. An ellipse has a major and minor axis, which are the longest and shortest distances across the center of the ellipse. These axes are perpendicular to each other (see section 1.1 page 2). Half of these axes, or the distances from center to edge of the ellipse, are called the **Semi-major** and **Semi-minor** axes respectively, with symbols **a** and **b**. The Semi-major axis is the value usually used to describe the overall size of an orbit.

Eccentricity - The foci of an ellipse are the points along the Semi-major axis such that the sum of the distances from the foci to any point on the ellipse is constant. An orbit of a small body around a more massive one will have the massive one located at one focus of the elliptical orbit. The **Focal length**, f, is the distance from a focus to the center of the ellipse. The shape of the orbit is measured by **Eccentricity**, e, which is defined as:

$$e = f/a$$

The higher the eccentricity, the narrower is the ellipse relative to the semi-major axis, and the greater the difference between the nearest and farthest points of the body from the one it is orbiting.

Periapsis and Apoapsis - The prefixes peri- and ap- refer to the nearest and farthest points of an orbit from the center of the body being orbited. Different suffixes are used to indicate what body is being orbited, such as perigee and apogee for the lowest and highest points of an Earth orbit, and perihelion and aphelion for distance from the Sun. The general symbols are q for perigee and Q for apogee, and can be found from the formulas:

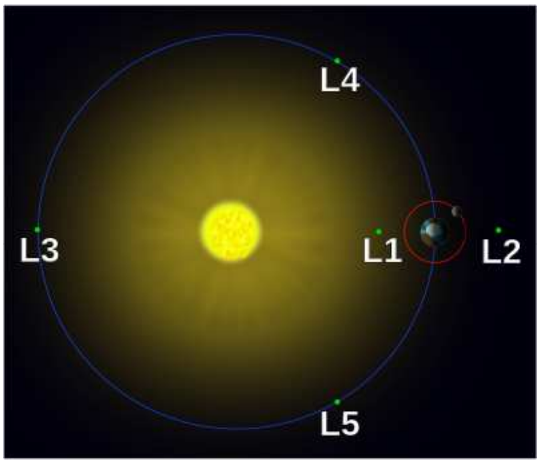
$$q = a - f = a(1 - e)$$

$$Q = a + f = a(1 + e)$$

$$f = ae$$

Lagrangian Points

Given two large bodies, such as the Sun and Jupiter, and a third small body, such as an asteroid, there are five points relative to the large bodies where the net forces keep the small body approximately in the same position relative to the two larger ones. They are called the **Lagrangian Points** after one of the mathematicians who discovered them. Three of these, called L1, L2, and L3, are unstable. If you move slightly away from the exact point, you will tend to move further away. The other two, L4 and L5, are stable. Slight movements around these points will not cause the small body to drift away, but rather orbit around the points. L1, L2, and L3 are located between, behind, and opposite the second of the large bodies, respectively (see figure). L4 and L5 are located in the same orbit as the second large body, 60 degrees ahead and behind it. As the largest planet, Jupiter has the largest collection of asteroids in it's Lagrange points. These asteroids are called the "Jupiter Trojans", since the first few at Jupiter were named after characters in the Trojan War. More generally, objects at the stable Trojan points of other planets are also called Trojans. Since most orbits are elliptical, the Lagrange points shift as the distance between the two major bodies changes.

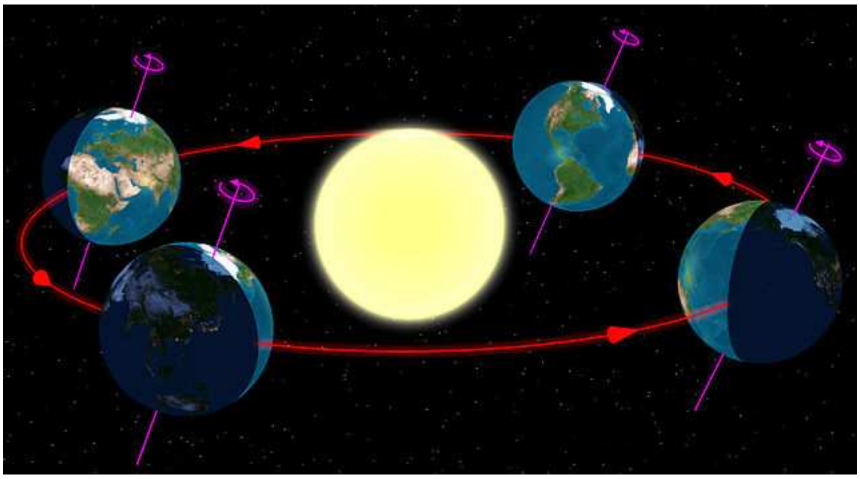


Rotation

Nearly every natural body in orbit also rotates, so that the direction from the body's center to a fixed surface point changes. This has several effects:

Rotation Period - This is the time it takes the body to complete one rotation with respect to the stars, the Sun, or the planet if it is a satellite of one. The most obvious effect of the rotation period is the day-night cycle on Earth. Some objects become locked into a rotational resonance with the parent body they orbit. This means the rotation period is a simple fraction of the orbit period. When the resonance is 1:1, it is called **Tidally Locked**, and the Moon is the most obvious example of that. The result is one side always faces the Earth, with a small wobble.

Axial Tilt - Rotation defines an axis of rotation. The places where the axis meets the surface of the body are called poles, and midpoint of the surface between the poles is called the **Equator**. On smaller bodies with irregular shape, the Equator may not be well defined. On larger bodies which are more or less round, the Equator has the largest distance from the rotation axis. **Axial Tilt** is the angle the body axis makes with the axis of the body's orbit. The rotational inertia of large bodies causes their rotation axis to remain relatively fixed relative to the stars. For example, the Earth's north pole points near the star Polaris, but this point is tilted 23.44 degrees from the axis of its orbit around the Sun. So it orbits, first one pole, then the other, points towards the Sun, causing seasonal changes (see figure)



Axial tilt of the Earth to the axis of its orbit, which is perpendicular to the orbit plane defined by the red circle.

Rotational Velocity - When you are on the surface of a rotating body, the circular motion about the axis produces an acceleration which can reduce gravity. The velocity and acceleration depend on the distance from the axis and the rotation period. For example, at the Earth's equator the rotation velocity is 465 m/s, which generates an acceleration of 0.0338 m/s^2, or about 3% of gravity. Thus the apparent weight is less at the equator than the poles. Large bodies, more than about 1000 km in diameter, have internal forces greater than the strength of the internal materials. Since rotation lowers gravity in some parts relative to others, the body flows into an ellipsoidal, or flattened, shape. This is called hydrostatic equilibrium. The rotation of any body lowers the difference between orbit velocity and surface velocity when they are in the same direction. In the case of Earth it is 5.9% less, but in the case of some asteroids, like 4 Vesta, it can be 36% lower. Very small objects which do not have structural flaws can even rotate faster than orbital velocity around them, producing regions where you cannot remain on the surface without mechanical aid.

Perturbations

Gravity forces extend to infinity. Therefore nearby large bodies, such as the Moon and Sun for the Earth, also add an acceleration component to the gravity of the Earth. This varies over time as their direction and distance changes. The side of the Earth facing the Moon, for example, is pulled by it's gravity 6.6% more strongly than the opposite side, because it is closer in distance. The difference in gravity between the near and far sides is called the **Tidal Force** because it is the source of ocean tides on Earth. Those tides happen because water is free to move towards the Moon, but continents are more restricted. Tides do distort the shape of the solid part of the Earth, but less than the oceans. Tidal forces affect other moons and planets too.

The lesser gravity of other bodies also affects the orbit of an object around the primary it is bound to. These are called **Perturbations**, because they perturb the orbit caused by the strongest gravity force. On long time scales perturbations can drastically affect an orbit. This is most obvious in the case of Jupiter and comets. Long period comets are often near escape velocity, so small velocity changes caused by the large mass of Jupiter can drastically change their orbits. This can lead to them becoming short period comets that stay near the Sun, or being ejected entirely from the Solar System.

Velocity Map

In space, physical distance does not matter as much as velocity, since space is mostly frictionless, and what costs you energy is changing velocity. This graph below shows the minimum ideal velocity relative to escape for the Sun's gravity well on the horizontal axis, and for planetary wells and some satellites and asteroids out to Jupiter on the vertical axis. There is no absolute reference frame against which to measure velocity. We choose escape as the zero point since it has the physical meaning of "to leave this gravity well, you must add this much velocity". Since you must add velocity to leave, the values are negative. If you have more than enough velocity to leave a gravity well, that is called **excess velocity**, and is measured infinitely far away

Determining Total Velocity

Total mission velocity is the sum of vertical and horizontal velocity changes on the graph, both in km/s. Note that the axes are different scales. To travel from Earth to Mars, for example, you first have to add velocity to climb out of the Earth's gravity well, add velocity to change orbit within the Sun's gravity well, then subtract velocity to go down Mars' gravity well. On the graph, that means taking the vertical axis velocity change from Earth surface to the top line, which is Solar System orbits (11.18), plus the horizontal segment to go from the Earth's orbit to Mars' orbit (2.3), plus the vertical change to go to the Martian surface (5.03). That gives a total mission velocity of 18.5 km/s, which has to be accounted for by various propulsion systems. To return to Earth, you then reverse the process.

The graph shows theoretical values (single impulse to escape). Real changes in velocity (delta-V) will be higher because (1) maneuvers are not perfectly efficient, (2) orbits are elliptical and inclined, and (3) propulsion systems are not perfectly efficient in performing a given maneuver. Various losses are measured by the difference between **ideal velocity**, the velocity you would reach in a vacuum with no gravity well present, and the actual velocity you reach in a given circumstance. So this chart is not an exact method for mission planning. It is intended to give a rough estimate as a starting point from which more detailed planning can start.

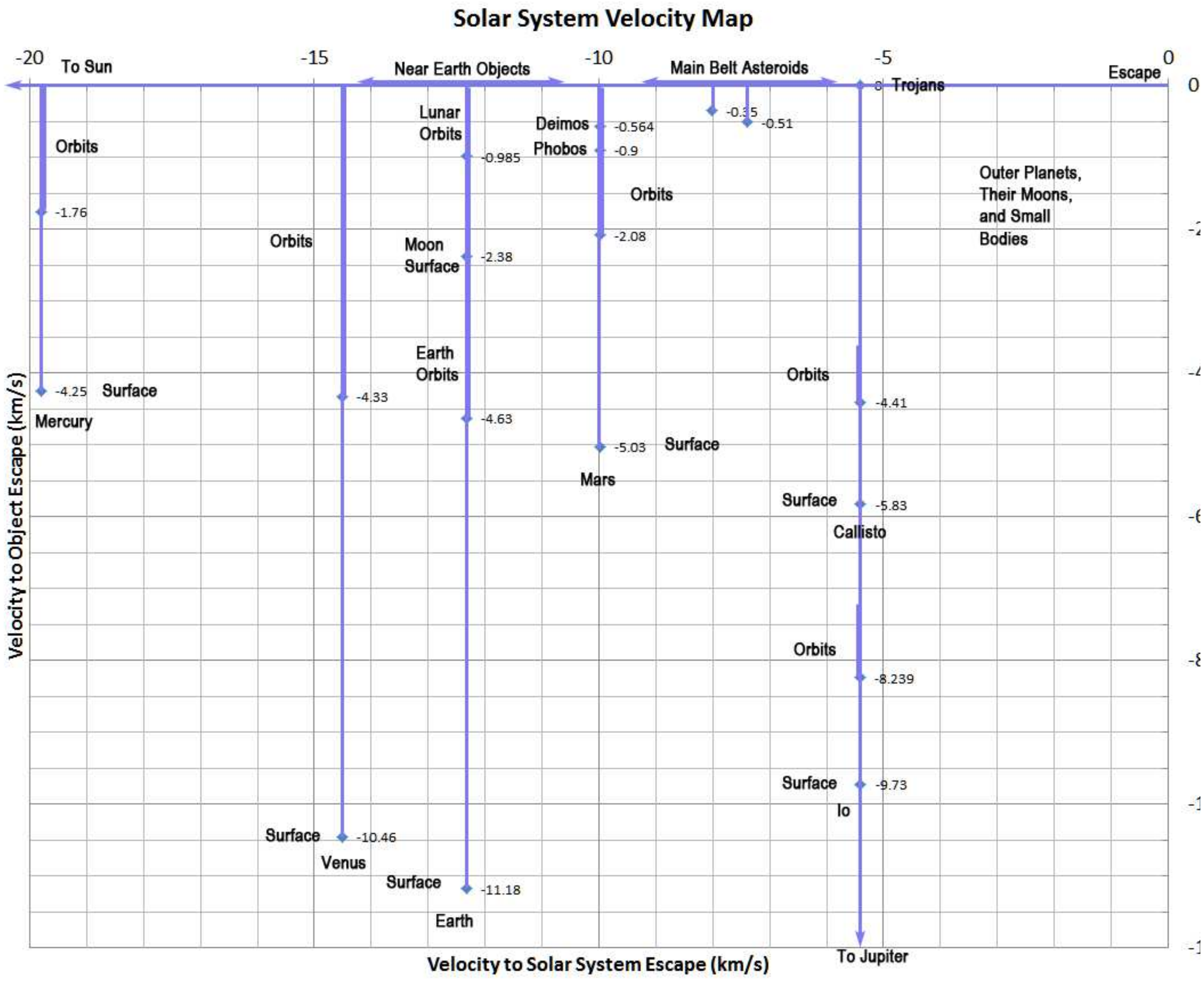
Velocity Bands

There are two velocity regions on the vertical axis for each planet or satellite. The lower **sub-orbital** region (thin purple line) is where you have enough velocity to get off the object, but not be in a stable orbit. Those orbits will intersect the body again. So they can be used to travel from point to point on the body's surface, but not to stay in motion for multiple orbits. The higher **orbital band**, shown with a thicker line, indicates enough velocity for a repeating orbit. The shape of the orbit matters too, but for circular orbits the lowest point in this band is an orbit just above the surface of the body, and the highest point is an orbit just fast enough to escape fom it's gravity well.

Since gravity varies as the inverse square of distance, relatively large velocity changes are needed for small altitude changes near the surface of a body. Conversely, near escape velocity, relatively small velocity changes can produce large changes in altitude, and at escape it produces a theoretically infinite change. In reality there are multiple gravity wells that overlap, so escape from Earth merely places you in the larger Solar gravity well, and escape from the Sun places you in the larger gravity well of the Galaxy

Solar Orbits

The top blue line represents orbits around the Sun away from local gravity wells. The surface of the two largest asteroids, 4 Vesta (-0.35) and 1 Ceres (-0.51) are marked, but the orbital bands for these two asteroids, and the entire gravity well for most smaller asteroids, are too small to show. Instead, the range of Solar velocities are shown for Near Earth Objects and the Main Belt between Mars and Jupiter. In reality the velocities of small objects in the Solar System are spread across the entire chart. The two marked ranges are just of particular interest. The surface of Jupiter and the Sun, and their sub-orbital ranges are off the scale of this chart because of their very deep gravity wells.



Powered Flight

Powered flight refers to trajectories and orbits which are *not* only under the influence of gravity and other natural forces. Rather they are the result of natural forces plus those created internally by an artificial system, or artificially applied to a system from the outside. The common example of internally created forces is a chemical rocket engine. Examples of outside forces are a powerful laser directed at a lightweight sail, or a gun that uses compressed gas to accelerate a projectile. The powered part of a flight may last a short time, as when a rocket launches into orbit. After that it coasts, only affected by gravity and other natural forces. A solar-electric engine, on the other hand, may operate over most or all of a flight or mission.

Ascent Trajectories

Circular orbit velocity at the earth's surface is 7910 meter/sec. At the equator the Earth rotates eastward at 465 meters/sec. So in theory a transportation system has to provide the difference, or 7445 meters/sec. The Earth's atmosphere causes losses that add to the theoretical velocity increment for many space transportation methods. The design problem is then to find the most efficient trajectory that minimizes losses.

In the case of chemical rockets, they normally fly straight up initially, so as to spend the least amount of time incurring aerodynamic drag. The vertical velocity thus achieved does not contribute to the circular orbit velocity, since they are perpendicular. So an optimized ascent trajectory rather quickly pitches down from vertical towards the horizontal. Just enough climb is used to clear the atmosphere and minimize aerodynamic drag.

The rocket consumes fuel to climb vertically and to overcome drag, so it would achieve a higher final velocity in a drag and gravity free environment. The velocity it would achieve under these conditions is called the 'ideal velocity'. It is this value that the propulsion system is designed to meet. The 'real velocity' is what the rocket actually has left after the drag and gravity effects. These are called drag losses and gee losses respectively. A real rocket has to provide about 9000 meters/sec to reach orbit, so the losses are about 1500 meters/sec, or a 20% penalty.

Boost From a Non-rotating Body

To go from a non-rotating body's surface to orbit requires that a rocket change its velocity from a rest velocity (zero) to a velocity that will keep the payload in orbit. If our rocket maintains a constant thrust during its ascent, then the total velocity change is

$$\int_0^{t_{orbit}} a\,dt = \int_0^{t_{orbit}} \frac{T}{m} - \frac{D}{m} - g\,dt$$

where a is the acceleration, D is the drag, and g is the planet's gravitational pull.

Boost From Rotating Body

[To be added]

Mass Ratio: Tsiolkovsky Rocket Equation

For any rocket which expels part of its mass at high velocity to provide acceleration, the total change in velocity delta-v can be found from the exhaust velocity v(e) and the initial and final masses m(0) and m(1) by

$$\Delta v = v_{\mathrm{e}} \ln \frac{m_0}{m_1}$$

The difference between the initial mass m(0) and the final mass m(1) represents the propellant or reaction mass used. The ratio of the initial and final masses is called the **Mass Ratio**. The final mass consists of the vehicle hardware plus cargo mass. If the cargo mass is set to zero, then a maximum delta-v is reached for the particular technology, and missions that require more than this are impossible.

Staging

A certain fraction of a vehicle's loaded initial mass will be the vehicle's own hardware. Therefore from the above rocket equation there is a maximum velocity it can reach even with zero payload. When the required mission velocity is near or above this point, dropping some of the empty vehicle hardware allows continued flight with a new mass ratio range based on the smaller hardware mass. This is known as **Staging**, and the components of the vehicle are numbered in order of last use as first stage, second stage, etc. Last use is mentioned because stages can operate in parallel, so the one to be dropped first gets the lower stage number.

The velocity to reach Earth orbit is approximately twice the exhaust velocity of the best liquid fuel mixes in use. So the rocket equation yields a mass ratio of e^2 or 7.39, and a final mass of 13.5%. This percentage is close to the hardware mass of typical designs, so staging has commonly been used with rockets going to Earth Orbit. We desire a rocket with a number of stages that optimizes the economic efficiency (cost per payload unit mass). The economic efficiency depends on a number of factors, the mass efficiency being only one factor.

Let us assume that we desire to launch a payload of weight *P*. The weight of each stage in the stack is

$$W_i = Pw_i$$

where w_i is a normalized weight for the stage.The total stack weight is thus

$$W = P\left(1 + \sum_{n=1}^{N} w_i\right)$$

The change of velocity per unit mass for each stage is

$$\Delta v_i = I_{sp_i} \ln \mu_i$$

where μ_i is the ratio of the weight before the burn of the *i*th stage to the weight after the burn of that stage. Thus, μ_i will always have a value greater than 1. The total change in velocity per unit mass for all the stages is then

$$\Delta v = \sum_{n=1}^{N} I_{sp_i} \ln \mu_i$$

Section 1.3 - Propulsive Forces

Space transport generally involves the application of forces to get to a desired location. The comprehensive listing of space transport methods found in Part 2 of this book can be categorized in two dimensions, as shown in the "Table of Space Transport Methods" at the start of that section. The X (horizontal) dimension is how the motive force is applied, and Y (vertical) dimension is the source of the energy used. (See also **Spacecraft Propulsion** at Wikipedia for another list of methods.) This section will discuss the forces, and section **1.4 Energy Sources** will discuss the energy sources. Transport methods can also be categorized by parameters like state of development, exhaust velocity, or efficiency. Section **2.11 - Comparisons** will consider these other parameters. We list all known forces considered for transport, without regard to practicality or state of development. How to select the best option for a given transport task will be covered later.

Propulsive forces can be divided into two large groups. The first group develop reaction forces from internal material which is expelled from the vehicle. The second group are forces generated by interacting with an entity outside the vehicle. The law of conservation of momentum (i.e the sum of changes in mass times velocity for the parts of a system is zero) requires that the force you impart to the object you want to move is matched by an equal and opposite force on something else. That something else can be mass you expel from the object, or an outside entity, but in either case the combined forces must be zero.

Reaction from Expelled Mass

Vehicles which use this type of force production are generally called **Rockets**, and the devices which expel the materials are generally called **Engines** or **Rocket Engines**. Specific types each have a descriptive or or common name noted later in the book where they are discussed. The forces generated by an expelled material are derived from **Newton's Laws**. Where F is force, *dm/dt* is amount of mass expelled per time, and v_e is exhaust velocity, the force can be found by

$$F = \frac{dm}{dt} v_e$$

For accelerating vehicles, we call the force **Thrust** and give it the symbol T From this equation we can see the possible ways to increase thrust are (1) increase the mass flow rate, (2) increase the exhaust velocity, or (3) some combination of both. The available material to expel, also known as **Reaction Mass** or **Propellant**, is finite when coming from an internal source. So it is usually preferable to increase velocity to get more total performance. Because this type of propulsion system is self-contained, it can operate in many environments, particularly the vacuum of space.

Exhaust Velocity

A parallel, unidirectional flow has all the mass expelled in a single direction. An expanding gas is generally emitted as a cone with some angular width (see Figure 1.3-1). Molecules not moving parallel the axis of the cone only contribute the parallel component of their motion parallel to the reaction force. This component is found from the cosine of the angle of motion times the molecular velocity. Molecules in an expanding gas also have a range of velocities determined by their temperature and how the flow was shaped. The average axial component over the whole flow is called the **Exhaust Velocity**. A derived unit is **Specific Impulse**. Where T is thrust, $\dot{m}_p$ is propellant flow rate, and g is standard Earth gravity it is defined as

$$I_{sp} = \frac{T}{\dot{m}_p g}$$

Figure 1.3-1 - Space Shuttle solid booster exhaust emitted as a cone of gas.

The units for specific impulse are in seconds, and are interpreted as how many seconds one unit of fuel can produce one unit of thrust at 1 gravity. As an example, the high energy propellant combination of $H_2 + O_2$ produces a specific impulse of about 450 seconds. Exhaust velocity in meters/second is is the preferred SI unit, since that is not Earth-centric by using the level of Earth gravity in the calculation.

In the list that follows, the approximate range of exhaust velocities is noted, and the list is generally in increasing order. Note that what we mainly use today (combustion gas) is among the lowest in performance. To achieve higher mission velocities than the characteristic exhaust velocity you can use large amounts of propellent, stack multiple stages of the same method, or use multiple different methods.

A. Bulk Solids (0-10 km/s)

Solid pellets or slugs are expelled via mechanical devices such as a rotary centrifuge, or an electromagnetic accelerator such as the **Mass Driver Reaction Engine**. The advantage is being able to use nearly any solid material as the reaction mass. One disadvantage is the relatively low exhaust velocity compared to Ion and Plasma engines. The extra work in extracting a suitable fuel for the latter types is usually much less than the gain from using 5-10 times less reaction mass. Another is the creation of a debris impact hazard by emitting large numbers of uncontrolled objects. A centrifuge or mass driver launching bulk mass from a body which is collected in orbit does not create the same impact hazard, even though it uses the same kind of devices. Mechanical devices are like centrifuges are limited to fairly low velocity, and have not generally been considered for space transport.

B. Microparticles (0 - 4 km/s)

In addition to using bulk solid, finely powdered solid microparticles or droplets such as from inkjet type devices may be accelerated by electrostatic forces after giving them an electrical charge. Advantages of this method are using unprocessed rock dust or single fluids, and enabling very small engines. Compared to bulk solids, microparticles or droplets are less of an impact hazard, though they may pose a contamination problem. Disadvantages include only working well in a vacuum or a very low density non-conducting medium. Any appreciable outside pressure would stop the microparticles by drag or collision. Relative to ion engines, the charge to mass ratio is lower, so the same electrostatic voltages result in lower exhaust velocity and less perfomance.

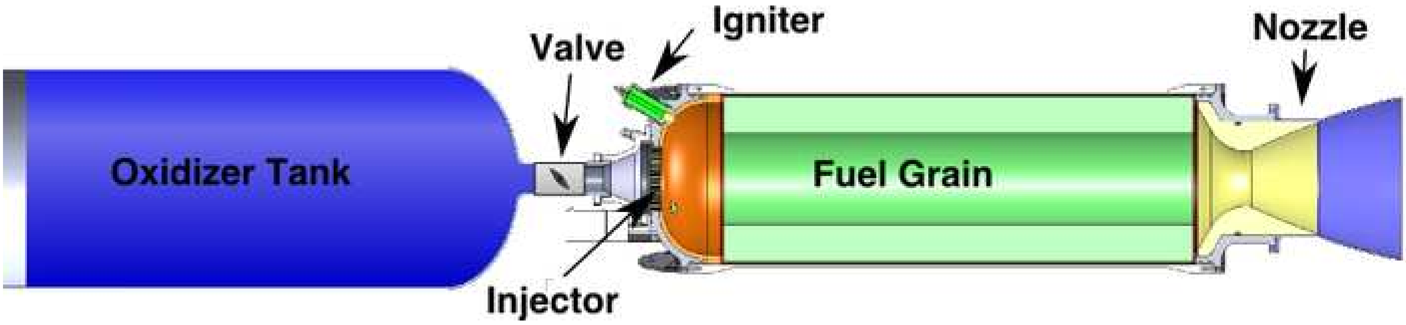
C. Gas Flow (0.1 - 10 km/s)

While in theory you can eject a liquid to obtain thrust, in almost all cases better performance can be gotten using a gas. This is due to the higher average molecule velocity and ability to extract energy from the gas expansion. So liquid is skipped among the solid-liquid-gas states of matter as a reaction method. However liquids are a useful as a way to store reaction mass due to higher density and lower vapor pressure. Higher pressures require more storage tank mass.

Gas flow includes ambient temperature gas such as the Nitrogen "cold gas" thrusters used in spacesuit maneuvering backpacks. Cold gas thrusters are useful when you don't want to damage hardware with a hot or chemically reactive exhaust plume, but they are very low performance (~0.5 km/s). Heating a low molecular weight gas (i.e. Hydrogen) allows much better performance due to the higher average molecule velocity. At sufficiently high temperature (around 3000K) the Hydrogen molecules will start to decompose to individual atoms, further increasing velocity. There are numerous possible methods of heating the gas, including electric discharge through the gas (arcjet), concentrated sunlight (solar-thermal), electric filament heaters (resistojet), or heat from a nuclear reactor (nuclear-thermal).

D. Combustion Gas (2 - 5 km/s)

In this method, hot gas is generated by chemical reactions in the propellant. For rockets the hot gas is expelled via a supersonic expansion exit nozzle. That type of nozzle produces the highest velocity in a narrow stream. Performance is limited by the reaction energy in the propellant, which is a maximum of about 15 MJ/kg for non-exotic fuel combinations. This provides a maximum of 5.5 km/s exhaust velocity in theory, and 4.5 km/s in practice. For atmospheric jet engines, some of the energy in the gas is used to drive a turbine and bypass fan blades, which greatly increases the affected mass flow. The remainder is expelled by a nozzle, but the nozzle is typically a simpler geometry. The use of external oxygen extracted from the air flow, and bypass air for mass flow, dramatically reduces the rate of fuel use, but also limits the flight velocity due to drag and heating.



2 - Hybrid rocket concept with liquid Oxidizer (blue) and solid fuel grain (green).

There are several types of combustion gas rockets. A **Monopropellant** has a single ingredient which is decomposed and heated by passing over a catalyst bed. A **Bi-Propellant** has two ingredients, a fuel and an oxidizer, which are generally mixed and burned in a **Combustion Chamber**. In a **Liquid Rocket** the two ingredients are stored in liquid form in separate propellant tanks, although one or both may be converted to gas before reaching the combustion chamber. In a **Hybrid Rocket** one of the ingredients is in solid form (usually the fuel), and the other in liquid form. In a **Solid Rocket** all the ingredients are in a finely mixed powder which has been cast into a solid form. A typical solid rocket formulation has an oxidizer like **Ammonium Perchlorate** (NH_4ClO_4), and a complex fuel containing powdered aluminum, rubber, and epoxy, which both binds allthe ingredients together and is part of the fuel being burned.

There are a large number of combinations of form and fuel mixtures for rockets, but only a few are used with any frequency. Because of the high thrust-to-mass ratio and ability to work in many environments, combustion gas systems have been by far the most popular for space transport so far, and almost exclusively the one for launch to Earth orbit. Their disadvantage is that the exhaust velocity of the best propellants is about half the necessary velocity, including various losses, to reach Earth orbit. Solving the Tsiolkovsky Rocket Equation (section 1.2), we find the mass ratio is at least 7.4, which leaves only 13.5% of liftoff mass for the vehicle and payload. The best propellant includes liquid Hydrogen as the fuel, which is very low density, so keeping the vehicle mass low enough for reasonable payload is difficult. If a denser and lower performance fuel is chosen, less structural mass is needed for propellant tanks, but the mass ratio is higher. This still leaves small payload mass. Until now, the solution has mainly been to use the rocket once, which allows lighter hardware, and discarding part of the hardware once some of the propellant has been used (staging). This is an expensive solution. For the future, a better approach is to move away from chemical rockets and their limited fuel energy in relation to what is needed to leave Earth. That will allow more of the vehicle mass to be hardware and payload, and less to be fuel. Designs can then be more robust and used many times, and the cargo can be a larger fraction of liftoff mass. The combination will drastically reduce costs.

E. Plasma (5 - 200 km/s)

In this category the propellant is heated to the point that the atoms disassociate into charged components (ions and electrons), then directed out of the **Plasma Engine** with magnetic fields. Heating can be accomplished by a vigorous electric discharge, intense microwaves, laser, or internal heating in a fusion plasma. Plasmas are hot enough to melt most materials, including the ones used for the rest of the engine, so they are usually contained by a magnetic field. Alternately the plasma density (and hence the thrust) can be kept low enough that the engine can disperse the heat gained. There is no theoretical limit to how hot a plasma can be. Since the kinetic energy of the exhaust increases as the square of the velocity, there is a practical limit based on how much energy you can supply. Plasmas can be physically confined by magnetic fields, but they also emit light that escapes to their surroundings. The surface of the Sun, for example, is a 5780 K plasma. So heating of the engine components will also limit a practical design. Many plasma confinement techniques derive from fusion research, which deals with extremely hot plasmas. Recent development of superconducting coils improves their efficiency, so design of effective plasma engines is fairly recent. Because their exhaust velocities can be ten or more times higher than for combustion gas, plasma engines are in active development, but have not reached operational use as of 2016.

F. Ions (2 - 200 km/s)

Ions are atoms from which one or more electrons have been removed. In an **Ion Engine** the propellant is first ionized using electron bombardment or RF oscillations (Figure 1.3-4). The positive ions, which represent nearly all the mass, are accelerated across a voltage gradient to high velocity. The voltage gradient can be a set of metal screens, or a charged plasma. To maintain overall charge balance across the vehicle, an electron gun separately emits negative charges, or the ions recombine in the charged plasma. Once outside the engine the combined emission is electrically neutral so that the vehicle does not accumulate a net charge. Compared to a plasma, an ion beam is generally lower density, so requires lower power levels to operate and generates lower thrust. The main difference is creating high ion velocity via electric forces rather than heating. Efficiency in converting power to thrust is similar (60-75%), as is maximum exhaust velocity. Like plasma engines, the energy required rises as the square of exhaust velocity. Practical limits on the power source then limits performance. Because ion engines can operate with less overhead in small sizes than plasma engines, and are derived from vacuum tube technology, they were developed earlier.

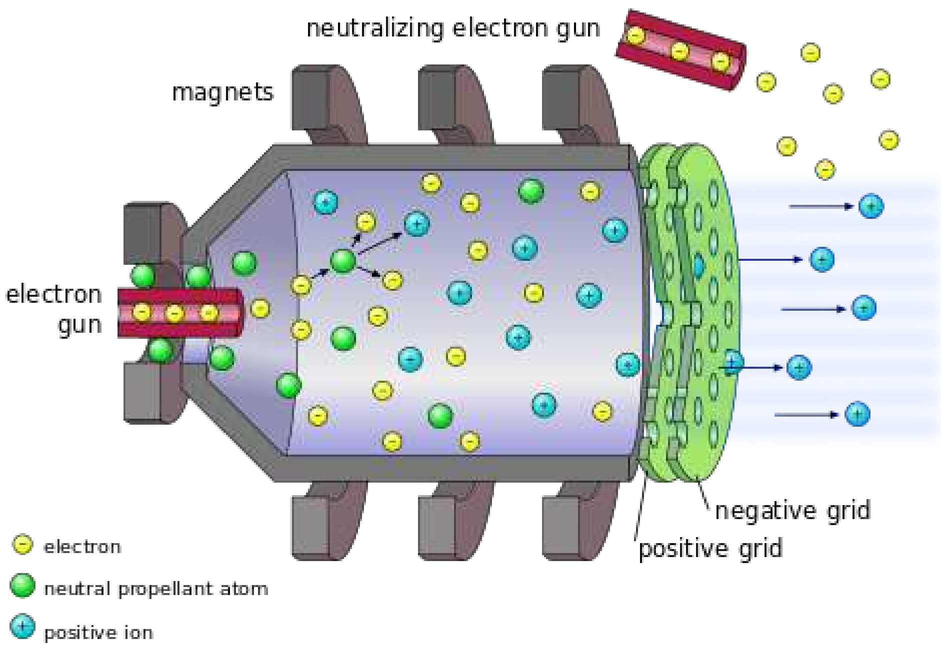


Figure 1.3-4 - Electrostatic ion engine.

Ion engines are in use, as of 2016, on many communications satellites, and a few planetary spacecraft where high total mission velocity is needed. In the future we expect their use to increase, along with the larger plasma engines, because they use an order of magnitude less propellant than chemical rockets. In addition, the last 20 years has seen the development of lightweight and efficient solar panels to provide power for them. Their disadvantage is low thrust, so they cannot be used on their own for launch or landing on large bodies. This can be overcome by combining them with other methods.

G. Atomic Particles (1000 km/s to near c)

Atomic particles include nuclei of atoms from which all electrons have been removed, single electrons, neutrons, or protons, or more exotic particles such as muons. Whereas an ion engine typically uses a single voltage gradient, a particle accelerator has multiple chambers that add successive amounts of energy to the particles. This enables exhaust velocities up to near the speed of light, *c* , or 299,792 km/s. Another method is direct emission of atomic particles from fission decay, fusion reactions, or antimatter decay. These have particle velocities that are a significant fraction of *c*. Charged particles (generally those besides neutrons) can be directed by magnetic or electric fields, while neutral particles must be allowed to leave in one direction and absorbed in the others to produce net thrust. Although accelerators on Earth have reached very close to c, which is the universal speed limit, no space mission has yet required this high a performance. So this reaction method remains theoretical at present.

In this category is the **Bussard Ramjet** concept, which is a kind of fusion rocket. Rather than carry all the fuel on-board, it uses a huge electromagnetic field as a ram scoop to collect and compress hydrogen from the interstellar medium.The field compresses the hydrogen until fusion occurs, then directs the energy as rocket exhaust, accelerating the vessel. The difference between the incoming interstellar gas pressure and outgoing exhaust provides the thrust. The Bussard Ramjet currently has conceptual difficulties and is very far from being used.

H. Photon Emission (1.0c exactly)

Photon momentum is calculated by E/c, where E is the energy and c is the speed of light. Thrust is then P/c where P is the power of the emission. Direct emission of photons, while low in thrust, has the highest possible exhaust velocity. For practical use, an extremely high energy source needs to be used, such as fusion or antimatter decay. Fairly simple thermal (black body) emission and reflector arrangements can align the light beam to produce useful thrust. A laser aligns the light very accurately, but the gain is small compared to a light beam with a width of a few degrees. The momentum contribution for an off-axis photon is cosine(a), where a is the off axis angle. For small angles that is very close to 1.0. Like atomic particles, no space transport mission has needed this high a performance, so it is a theoretical method at present.

Unbalanced photon emission as an incidental effect has been noted for the **Pioneer** spacecraft and with respect to small asteroids, where it is called the **Yarkovsky Effect**. However these accelerations are very small and not intentional.For useful propulsion a very large amount of photon energy needs to be emitted in a chosen direction.

Discussion

Split Energy Source:

Many of the above listed propulsion methods have separate propellant mass and energy to accelerate that mass. With these methods there is a design trade between mass and energy. The equation above gives force as mass flow rate times exhaust velocity. To get a desired mission velocity change, or ΔV, you can minimize the mass used, and use a lot of energy to generate a high exhaust velocity. Alternately you can minimize energy and accelerate a larger propellant mass. It is a trade-off where using more of one requires less of the other. The kinetic energy required for a given exhaust velocity is KE = $mv^2/2$, so it increases faster than the mass flow is reduced. Typically engine efficiency is relatively constant in the normal operating range. So source power level, which is KE/efficiency, also tends to increase as the square of exhaust velocity. Higher power levels require more power source mass, and represent overhead beyond the cargo you are delivering. Whether that mass increases linearly with power level or not will affect the optimum exhaust velocity. In addition, the choice of completing the mission in the least time, least propellant used, or somewhere between those extremes will affect the optimum exhaust velocity to use.

Once a design has been set for a vehicle, there is additional optimization possible within a particular mission. The design and mission fix the quantities of propellant mass and mission ΔV. If the engine type allows variable exhaust velocity, it takes less total energy to "dump" the early exhaust mass at a lower exhaust velocity V_e, and then eject the later mass with a higher V_e, than to eject at a constant V_e. If the goal is minimum time and you have a fixed energy source, this will be the optimum operating profile.

For missions which use solar power as a source, the optimal operation is more complex. First, solar flux varies as the inverse square of distance from the Sun. Second, solar powered orbit changes around a large body introduces shadowing and a stronger gravity field, which takes longer to traverse. For low thrust engines, short thrust intervals require less ΔV by up to a factor of $\sqrt{2}$, but close orbits are in shadow up to half the time and limit when you can generate thrust. When the local gravity acceleration is low relative to the vehicle acceleration, which happens farther from the body, then short thrust intervals do not impose a large time penalty. Deeper in a gravity well, they do. Lastly, for bodies with radiation belts, like Earth and Jupiter, time spent climbing through those belts can damage solar arrays, other hardware, and human passengers. In complex cases like these, the best thrust plan is found using a numerical simulation which divides the mission into small time steps and adjusts the variables looking for the best result for whatever parameters the mission planner wants to optimize (time, fuel used, radiation exposure, or others).

Combined Energy Source:

Other propulsion methods listed above, such as combustion gas rockets, have propellants which provide both the mass flow and energy supply to generate an exhaust velocity, V_e. In these cases the Tsiolkovsky rocket equation shows that higher velocity is always better. Combustion engines have been developed with extraordinary power/mass ratios, up to 2.7 MW/kg. This enables them to lift not only themselves, but the entire vehicle's mass against the Earth's gravity. This makes possible launch to orbit without external assistance. Their performance is limited, however, by the available energy in the fuel, which in the case of Earth is about half that needed to reach orbit. Atmospheric engines obtain much of the mass flow and energy from outside the vehicle, so they bypass the limits of a purely internal propellant. This usually comes at the expense of increased mass from items like inlets, turbines, and wings. This lowers the power/mass ratio, and of course only functions while you are still in the atmosphere. The choice to use atmospheric engines will therefore depend on many factors beyond propellant energy. Atmospheric engines can have as much as 10-20x lower fuel use for a given thrust, so that is a strong incentive despite the complications.

Comparing split vs combined energy sources, a separate energy source relieves the limits on energy per propellant mass used. This allows higher exhaust velocities and lower propellant use, often by an order of magnitude or more. The drawback is current technology for these methods usually generates much less than one Earth gravity, because energy sources like solar panels are below 200 W/kg, far below combustion engines. This limits their use for the important job of getting from the ground to Earth orbit. Once in orbit, their performance advantage will often make them the preferred option. This is because the best chemical propellants contain about 15 MJ/kg of energy. A 175 W/kg space solar panel (state of the art as of 2016) will produce that much energy in about a day. Since the panel typically lasts 15 years, it will produce about 5000 times more total energy over its life.

External Interaction

The second major group of propulsive forces are applied by or against some external object or field, rather than expelling some material from the vehicle. Since no reaction mass is consumed, the Tsiolkovsky rocket equation does not apply. We list them roughly in order of performance, but in this group the measure is effective velocity change rather than exhaust velocity.

I. Mechanical Traction (5 km/s)

This method uses mechanical forces applied to a vehicle or payload from an external source. Applications of this method include:

- Using a cable or net to capture and decelerate the vehicle relative to a destination.
- An elevator to climb a tower or cable at a constant velocity
- A tow cable between two vehicles or between a fixed installation and a vehicle, for acceleration or lift.
- A rotating structure or cable to provide radial or angular acceleration.

Mechanical devices transmit forces by atomic bonds in the materials used. They are limited by the strength of the bonds, and so the bulk material strength, to around 3-5 km/s for existing materials. The theoretical strength limits for carbon nanotubes (100 GPa tension, per Zhou, **Ultimate Strength of Carbon Nanotubes**, Phys. Rev. B, v 65 p144105, 2002) and diamond (90 GPa compression, per Telling **Theoretical Strength and Cleavage of Diamond** Phys. Rev. Lett. v84 n22 p5160, 2000), given a density of 3500 kg/m^3 and a design margin of 2.8, produce a theoretical velocity of around 12 km/s. However, real materials accumulate defects, even if at an atomic scale. So the usable strength in practice is much lower than the theoretical value. A large advantage for mechanical systems is their inherent ability for repetitive use. This divides the initial cost by the number of uses over the economic life of the system.

J. Friction (2 km/s)

This method uses frictional forces against a solid surface to lower relative velocity. If your intent is to slow down on, for example, the Moon, you can have a flat paved runway or a raised rail, and simply use mechanical braking against the runway or rail to stop. To accelerate to orbit, you can grasp a trailing cable or rail on an orbiting platform and apply friction to gain velocity. This method has the virtue of being simple. Space velocities, however, often represent more kinetic energy than it takes to heat and melt materials. So the friction forces must be distributed in time, rate, or total amount to prevent overheating. It is therefore more useful for planets or satellites with low orbital velocities, and for small velocity changes as part of a larger total.

K. Gas Pressure (6 km/s)

This method uses external gas pressure differences to apply forces to a vehicle. It includes all types of guns, where the gas is confined in a tube to accelerate linearly. Gas expansion is limited by the temperature and molecular weight of the gas. Hot Hydrogen then yields the best performance, which is limited to around 9 km/s in the best case, and about 6 km/s for practical designs. There are a variety of methods to generate hot gases. The oldest, conventional firearms, depend on fast combustion of a solid propellant grain. Newer versions function via combustors , fuel-air detonation, particle bed heaters, or other methods to create heat and pressure in a short time. When the hot gas so created is not sufficient, a two-stage method transfers the energy by means of a piston to a lighter gas (usually Hydrogen). A properly tuned piston mass can result in higher temperatures and pressures in the light gas than in the first stage.

Various kinds of guns have been used for high velocity research since about the 1960's. Most have been used indoors, but at least two have seen outdoor use. They have not been used for space launch yet, but that is a matter of size and location, rather than technology level. An extreme use of gas pressure is to use a fission or fusion device to heat a large amount of gas to plasma temperatures in a chamber, which then is applied to accelerate a vehicle. This concept remains theoretical because of bans on testing and the tendency to destroy the chamber.

L. Aerodynamic Forces (4 km/s)

These forces include lift, buoyancy, and drag. Wings and fan blades develop lift by pressure difference across the upper and lower surfaces when a fluid such as air flows across them. At higher velocities, a **WaveRider** type inclined surface rides the lower shock wave it generates to create lift. Buoyancy develops lift by having a lower density than the surrounding fluid, such as in a balloon. Drag generates forces by accelerating the surrounding fluid in the direction you are moving, thus producing a force opposite your motion. A parachute, for example, is shaped to capture and accelerate the maximum amount of air, producing the maximum amount of drag to slow you down. Lifting forces invariably create drag, and tend to create more at higher velocities. Heating also becomes a limiting factor at high velocity, tending to set a limit of about 4 km/s for aerodynamic forces, except for re-entry systems where the heating is efficiently dissipated. Since aerodynamic forces in themselves always create friction and other dissipation, by themselves they can only slow a vehicle relative to the fluid. They require a source of thrust, like a jet engine, to accelerate.

M. Photon Reflection (185 km/s)

Instead of an internal source, as in H. Photon Emission above, this method uses an external source of photons, such as a star or laser. Normal (perpendicular) reflection produces a force of

$$F = \frac{E(1+R)}{c}$$

where E is the incident energy on the reflector, R is the reflectivity (fraction of incident light reflected), and c is the speed of light. This is nearly twice the force from emission, because the photon changes velocity by twice the speed of light from forward to backward. It is slightly less than twice because no reflector is 100% efficient.When the reflection is not perpendicular, the force is reduced by off-axis cosine losses. The quantity of light (photons) is not limited by an internal power source, so can reach relatively higher forces than emission, but still generally low compared to other methods. A solar or light sail uses thin but very large reflectors in order to intercept the maximum amount of light. To get the highest accelerations, the maximum ratio of reflected energy to mass is desired. Tungsten may be a dense element, but can operate at much higher temperature close to the Sun where the light intensity is higher. Where maximum temperature is not required, a light alloy like Magnesium-Aluminum, with high reflectivity, is preferred.

Given a sufficiently powerful light source, quite high velocities are possible. Using natural light sources such as the Sun we get approximately 8.2 microNewtons/m^2 at the Earth's distance. The gravitational force from the Sun on a lightweight sail may be 0.35 times this. The **Lightness Ratio**, LR, or ratio of light pressure to gravity then determines the maximum escape velocity by

$$V_{max} = (LR - 1)V_e$$

where V(e) is the local escape velocity. So paradoxically, to reach maximum final velocity you want to start as close to the Sun as possible, where local escape velocity is higher. For a lightweight sail, this might be 0.2 AU, limited by the melting point, and an escape velocity of 100 km/s. Thus the maximum final velocity is 185 km/s. Any velocity less than this maximum is possible. The practical limits using artificial light sources such as a laser are unknown, since no lasers with high enough sustained power to be useful for this purpose exist. In theory a powerful enough laser can accelerate a sail to substantial fractions of the speed of light, before accounting for drag from the interstellar medium.

By tilting a sail from perpendicular to the light source, off-axis forces can be generated. The net force will be approximately perpendicular to the sail and allow inward spiral orbits by directing the force against the orbital velocity, or tilting of the orbit plane by directing the reflection perpendicular to velocity. Sub-orbital or zero velocity motion can be produced by balancing light pressure against local gravity. These are not orbits because they depend on constant forces. Rather they fall into the class of powered trajectories.

N. Particle Deflection (100 km/s)

The thin plasma and gas emitted by the Sun's heated outer layers is called the **Solar wind**. The wind has typical velocities of 400-750 km/s relative to the Sun, and extends outwards to about 125 AU, where it encounters the interstellar medium. The force per area due to light pressure is much greater than that due to solar wind flux. A **Magnetic Sail** is a proposed method of spacecraft propulsion. It would use a magnetic field to deflect charged particles in the plasma wind. Since the field itself is non-material, in theory it can be large enough to get useful thrust, even though the area density of the solar wind is very low. By tilting the field, a sideways force can be generated. This method is limited to the velocity of the wind, and generally to producing forces away from the Sun. An artificial beam of particles could also, in theory, be used to apply a force to a vehicle, which either absorbs or deflects the beam. This remains theoretical because powerful enough beams are not available, and keeping them focused over typical distances in space is difficult.

O. Magnetic Field (20 km/s)

This method generates a force by using a current carrying wire, coil, or magnet to react against other magnetic fields (natural or man-made). In this category fall magnetic levitation or **Maglev**; **Coilguns**, which use a series of timed coils; **Railguns**, which use two high current rails and a plasma short across them; and **Electrodynamic Propulsion**, which reacts against a natural magnetic field. It can be used both for net thrust and drag, or for torque forces to rotate a vehicle. Magnetic flywheels are often used to orient or rotate spacecraft. The Sun and some planets have a natural magnetic field to react against. Many artificial satellites also use the magnetic field to rotate the satellite to a desired orientation using a **Magnetorquer**. For example, see Galysh et al, **[1]**. In theory there is no limit to the velocity you can reach using magnetic fields. In practice, the natural fields such as the strong one around Jupiter, or the practical scale or field strength of an artificial accelerator, limit velocity changes to around 20 km/s.

P. Gravity Field (20 km/s)

Gravity forces accelerate objects towards any nearby mass. Normally this results in an orbit or simply falling towards an object. A **Gravity Assist** is purposely choosing a hyperbolic path to change the direction, but not the total amount, of your velocity relative to a given object. The maximum change is twice the escape velocity of the object, when the direction is changed 180 degrees. To reach a desired mission destination, usually much less than this can be done. Since a planet, for example, is moving with respect to the Sun, changing direction relative to the planet can change the total velocity relative to the Sun. By conservation of momentum, the planet also must change velocity, but since it is much more massive, the velocity change is small enough to ignore in most cases. Within the Solar System, the achievable velocity changes are about 20 km/s, and may need multiple gravity assists to reach this level. The major advantage of gravity assist is it does not need internal propulsion except to line up the flyby, and therefore saves propellant mass. A disadvantage is one or more gravity assists take extra time to complete, and restrict the choice of trajectories because you can't choose where the moon or planet you are using will be at a given time.

Gravity fields extend to infinity, so every place in the Universe has a field, and almost everywhere has a non-zero net field. So all space transport has to account for gravity forces in mission planning. Usually it is an obstacle to overcome, in reaching orbit or changing orbits, but in some cases, such as gravity assist, it can be used to advantage. Gravity changes as the inverse square of distance from an object. If your vehicle or object is elongated, the **Gravity Gradient**, or difference in gravity between the lower and upper ends, can also be used as a torque to stabilize your orientation. These gravity gradients exist whether or not you use them, so they must also be taken into account as a force to be overcome in system design.

Discussion

Compared to the expelled mass group, external interactions do not consume a finite supply of reaction mass. So these methods would be preferred, all other factors being equal. Reaction mass, though, can be expelled under a wider range of circumstances in most cases, and usually with a higher thrust/mass ratio. So it is not possible to state a general conclusion as to which approach is better. The choice would depend on a variety of detailed circumstances, including destination, trajectory, desired trip time, cargo mass, frequency of trips, and how far in the future the transport occurs. The latter affects what technologies are available and how ready for use they are. One approach to selection is to look at the external interaction group first, to see if any can be applied to the job at hand, then look at the group that uses up reaction mass second, since that requires overhead above whatever cargo you are trying to deliver.

Section 1.4 - Energy Sources

Energy in General

All of civilization requires energy to function, including about 8 MJ/day (~2000 Calories) of food energy per person. Space systems are no exception, They require energy both for propulsion and for other systems like life support, computers, and materials processing. The general field of energy is vast, involving many fields of engineering. Energy for space projects involves an equally wide range. So we can only provide an introduction, and supply references for further study. In this section we survey the range of possible energy sources for all types of space systems. Since the book is oriented to future space projects, we list a number of sources that are not developed yet, but are possible according to known physics. These sources can then be used for the propulsion methods tabulated in Part 2, and for other engineering purposes described in Part 3 and later. By making a two dimensional table of energy sources vs propulsive forces, we can categorize all possible propulsion methods, and we do this at the start of Part 2 of the book. We have not yet developed a similar table to neatly categorize the other systems in a space project.

Proximate vs Ultimate Energy

The principle of **Conservation of Energy** states that energy can neither be created nor destroyed, merely transformed from one kind to another. So all energy for a project must come from a preexisting source. For a given project you can distinguish a **proximate** energy source, which is in the form consumed by the project, and an **ultimate** energy source, which traces the proximate source back through previous transformations to it's original form. Ultimately, all energy traces back to the creation of the Universe, but for engineering purposes we rarely go back that far, and are generally concerned with proximate sources.

For current space systems the energy is typically stored internally as chemical energy in the case of launch vehicles, or uses sunlight in the case of satellites. In the future, energy needs are likely to change, and the sources of energy will also change. Permanent locations, such as a large orbiting habitat or surface base, generally need continuous sources of energy to operate. Devices like batteries become unwieldy at that scale to supply power for the night part of Earth orbit or the two week Lunar night. Future projects may also need much greater power levels for tasks like processing of local materials. So the following headings attempt to include all potential energy sources, including many not yet used, but which may become useful in the future space projects. We list them all so that designers know the full range of possibilities, from which they can then select viable options for a given task. We exclude sources such as human and animal power from consideration here, both due to their low power levels, and because living things are not subject to the same kind of engineering design as we apply to non-living system. We also exclude sources like wind and geothermal, which mostly apply to Earth. Last, we includes some energy storage methods, which are not strictly sources. Energy storage, however, is often a necessary and important part of system design.

Energy References

A starting point for understanding energy sources in general, not just as they apply to space projects, is a National Academies book **America's Energy Future**, 2009. About 150 other books are available for free download in the **Energy and Conservation** topic from the same site. Encyclopedic references on the topic of energy, include the **Encyclopedia of Energy**, **Encyclopedia of Energy Engineering and Technology**, and the **Macmillan Encyclopedia of Energy**. These are often expensive reference books, so library or other sources are recommended to get access. There are numerous engineering books on more specific aspects of energy systems. Wikipedia also has an **Outline of Energy article**, with many links. Some of the concepts listed below are currently theoretical, so they are not well covered in reference books about current energy use or engineering. Information about them will mostly be found in research reports and scientific/technical papers.

Mechanical Sources

Mechanical Energy includes energy stored by virtue of previous work, as in compressed gases, and that which exists by virtue of position (potential) and motion (kinetic). Objects in orbital motion have a combination of potential and kinetic energy.

A. Compressed Gas

Although **Pressure vessels** are strictly an energy storage method, for space missions the tanks are normally pre-filled. So they operate as a proximate energy source in flight. The available energy, W, stored in a pressurized tank, can be found from

$$W = p_B v_B \ln \frac{p_A}{p_B}$$

where B represents the high pressure and A represents the low pressure, and p and v are pressure and volume respectively. So a 1 cubic meter tank with a high pressure of 20 MPa and a low pressure of 10 Mpa would provide 13.8 MJ of available energy. Compressed gas is a low density energy storage method. It is often used in space vehicles for tasks like cold gas thrusters and pressurizing liquid fuel tanks. Its chief advantages are simplicity, requiring just a storage tank and a valve, and rapid release of the stored energy. When larger total amounts of energy are needed, a higher density but more complex system is often preferred.

B. Potential Energy

Potential Energy is the ability of a system to do work by virtue of it's position or configuration. In space projects this is usually position relative to the gravity well of a massive object such as a planet. A simple hypothetical example is a stationary space elevator cable. While raising a cargo, electricity is converted to potential energy of height. When lowering a cargo, the potential energy can be extracted back to electricity. The formula for potential energy U was given in **Section 1.1 - Physics** as

$$U = -G\frac{m_1 M_2}{r}$$

The difference in energy at two radii gives the amount of potential energy stored or released over the distance. For small changes in radius (height) relative to the distance r, the potential difference can be approximated by the average gravitational force (weight) times height. On planetary surfaces, large amounts of available mass can be used to store potential energy. On Earth this is done with dams for hydroelectric power. On other bodies, a mountain and a pile of rocks can serve the same purpose. Transporting the rocks up and down the mountain can serve to store or release energy.

C. Kinetic Energy

Kinetic Energy is that which an object possesses by virtue of it's motion. Its formula was given in **Section 1.1 - Physics** as

$$KE = \tfrac{1}{2}mv^2 = Fd$$

where KE is the kinetic energy, m is the object's mass, and v is the velocity. It is also equal to an accelerating force *F* times the distance *d* it is applied over. An object in orbit has both kinetic energy in it's orbital velocity and potential energy in it's altitude. In an elliptical orbit, it continuously exchanges altitude for velocity. So it also exchanges potential and kinetic energy, but the combined total stays the same.

Rotating objects such as a space station or reaction flywheel have a form of kinetic energy in it's motion around an axis. Rotational energy is E(k) calculated by

$$E_k = \frac{1}{2}I\omega^2$$

Where ω is the **Angular Velocity** in units of radians/second, and I is the **Moment of Inertia** of the mass about the center of rotation. The moment of inertia is the measure of resistance to **Torque**, or rotational force, applied on a spinning object. The higher the moment of inertia, the slower it will spin when a given force is applied. Moment of inertia depends on the distribution of mass in the rotating object. The farther out a given portion of the mass is, the larger the contribution. Formulas for many shapes are found in the **List of Moments of Inertia**. For complex shapes, the total moment can be found by dividing it into simpler parts and summing the individual moments.

Some examples of moment of inertia fomulas are:

$$I = \frac{1}{2}mr^2 \text{ for a solid cylinder,}$$

$I = mr^2$ for a thin-walled empty cylinder, and

$$I = \frac{1}{2}m({r_{external}}^2 + {r_{internal}}^2)$$ for a thick-walled empty cylinder

Kinetic energy can be exchanged for other forms of energy by gravitational forces, as in gravity assist maneuvers; or electromagnetic forces, as in many electromechanical devices. Potential energy can also be exchanged for kinetic energy via the **Oberth effect** by expending propellant deep in a gravity well.

Chemical Sources

Chemical sources are an arrangement of atoms in a higher energy state which are converted to a lower energy state by a **Chemical Reaction**, releasing the difference. Combustion is the most common way to release this energy, accounting for over 80% of total human energy use. Batteries are characterized by a reversible reaction, so that the same device can store and release energy multiple times.

D. Fuel-Atmosphere Combustion

The Earth's atmosphere, neglecting the variable amounts of water vapor, contains 20.95% Oxygen (O_2) molecules, which react with many other compounds to release energy. This Oxygen is the byproduct of **Photosynthesis** in living things. In the case of aircraft, a hydrocarbon fuel such as **Kerosene** is reacted with the atmospheric Oxygen in an engine. Since only the fuel is carried internally to the vehicle, the energy released, about 43 MJ/kg, is about three times as much as when both ingredients are carried internally, such as in a typical rocket. Large amounts of Oxygen in an atmosphere is unstable, because it is so chemically reactive. It only exists on Earth in this form because plants constantly produce it. So this energy source is not available on other bodies. The reverse option is available on a body such as Titan, which has a hydrocarbon atmosphere. In that case, Oxygen can be the carried ingredient, and burned with the surrounding atmosphere. For atmospheres which are mostly CO_2 (Venus and Mars), which is an end product of combustion, or bodies with no atmosphere at all, this energy source is not available.

E. Fuel-Oxidizer Combustion

The energy source in conventional rockets is **Combustion**, where both the fuel and oxidizer are supplied from internal sources. The ingredients with the highest reaction energy, Hydrogen and Oxygen, provide 15 MJ/kg of propellant. Although lower in **Specific Energy** than D. Fuel-Atmosphere Combustion, it is not restricted to operating in the Earth's atmosphere. Liquid rocket engines also have extraordinary power-to-mass ratios. This enables launch trajectories from large bodies like planets. Combustion can also be used as a secondary power source in **Auxiliary Power Units**. Because the rate of energy release is very high, combustion is useful when high power levels are needed. The efficiency of combustion engines is typically 1/3 to 2/3, so other options may be preferred when that is an important factor.

F. Chemical Battery

An **Electric Battery** is a device which converts stored chemical energy to electricity, and in a storage battery also reverses the reaction. Common examples of chemical batteries include the **Lead-Acid** type used in automobiles, and the **Lithium-Ion** type used in many portable devices. Depending on battery type, they generally store less then 1 MJ/kg, considerably less than combustion. The ability to cycle energy in and out multiple times can outweigh the lower energy density. An example is the International Space Station, where large batteries supply power in the shadowed part of its orbit.

A **Fuel Cell** is a type of battery where the reactants are stored in external tanks, rather than in a sealed battery case. It can have high specific energy because the tanks are lightweight compared to electrolyte solutions. Fuel cells have therefore been used in space projects, such as the Space Shuttle Orbiter. A Hydrogen-Oxygen fuel cell combined with an **Electrolysis** unit to convert the resulting water back to Hydrogen and Oxygen can supply energy storage with a specific in the range of 3-10 MJ/kg. Sealed batteries are simple and reliable, and can be made in very small sizes. Fuel cells have higher specific energies, but are more complicated devices, since they need valves and a way to pump and store the various chemicals.

Thermal Sources

Thermal Energy is the internal energy of a system due to its temperature. It comes from the kinetic and vibration energies, and the attractive potentials of the molecules or other particles making up the system. Thermal energy can be stored for later use, or added from an outside energy source for immediate use. We sense high temperatures as heat, and even higher temperatures as visible light. Energy naturally flows from higher to lower temperature areas by conduction, radiation, and convection. When a temperature difference exists, some of the thermal energy can be converted to other forms and used. An example is a steam turbine that generates electricity from the difference between hot steam and the cooled outlet.

G. Thermal Storage Bed

For locations like the Lunar surface, which has a long night, solar power is not effective half the time. So storing heat in a **Thermal Energy Storage** system may be a viable option. Heat is put into rocks during the daytime and extracted from them at night to run a generator. The rock bed is enclosed in a container, and gas transfers heat to a turbine for generation, and from a solar collector for storage. Since the rock can be obtained locally, the energy stored per mass of installed equipment is fairly high. Environment temperatures during the Lunar night are quite low, and this can be enhanced by thermal shields between a radiator and the ground and daytime Sun. So the temperature difference between the storage and rejection temperature, and therefore efficiency, can be fairly high.

Some bodies, like Earth and Jupiter's moon **Io**, have relatively high interior temperatures. They serve as a natural thermal storage bed by the low thermal conductivity of surface rock. The source of heat can be radioactive decay or tides. This energy can be put to use by drilling down to high temperature regions, and exploiting the difference between those and surface temperatures.

H. Concentrated Light

A number of industrial tasks require heating, which is easily done in space by concentrating sunlight. Examples are heating of raw materials to extract volatiles, or maintaining temperature and growing ability in a Mars greenhouse. The concentration ratio determines the maximum **Black Body** temperature that can be reached, up to the temperature of the light source. In the case of the Sun, the upper limit is the Sun's surface temperature, 5,775 K, less reflection losses and radiation losses from the object you are heating. Since Tantalum-Hafnium-Carbide, the highest melting point substance known, melts at 4200 K, concentrated sunlight should be sufficient for most industrial processes.

For space transport, a reaction mass can also be heated by concentrated solar or artificial light. Lighter molecules can be used than the exhaust products of chemical reactions, so higher performance can be reached. Lack of powerful enough lasers limits their use for propulsion at present, but sunlight is widely available in space.

Electrical Sources

Electricity is the set of phenomena associated with the presence and flow of electric charge. Common examples are **Electric Current**, in the form of electrons moving in a conducting wire, and **Lightning**, a powerful electrostatic discharge through a plasma channel in a storm. Electricity is a very versatile energy source because it can be converted to other forms efficiently, controlled in both tiny and large amounts, and moved about from place to place relatively easily. There are a number of ways to produce and distribute electrical energy.

I. Power Line

Most space projects use electrical energy in some form. The parts of a project on Earth are often by far the largest. These include factories to produce the vehicles and spacecraft, launch sites, and control centers. Typically they get their electricity from a network of **Electric Power Transmission** and **Distribution** lines. This is distributed to the point of use by local **Electrical Wiring**. What distinguishes these three is the scale of power, P, in Joules/second, or **Watts**, and the voltage and distance the power is moved. Most wires have electrical **Resistance**, R, which is a measure of the difficulty in carrying a current, I, using a voltage V They are related by the formulas

$$V = IR \qquad and \qquad P = VI$$

The resistance causes some of the energy to be converted to heat. The amount of power converted, P, is found by

$$P = I^2 R$$

When efficient delivery of energy, and not heating, is the intended purpose, you want to minimize the resistance heating. Since resistance is a material property, by this formula you want to use a low current,

I. By the previous formula, the useful power P = VI, so a low current implies a high voltage. Therefore long distance lines are operated at higher voltages, and voltage changes are provided by **Transformers** as needed.

The same principles will apply to the space portion of a project when the distance between the point of generation and the point of use is large. Additionally, mass is usually a factor for space systems. So besides minimizing losses from resistance and transformer efficiency, you want to minimize the mass of the wires. The International Space Station is an example where generation and use are separated by an average of about 50 meters. This is because the Station is intended as a zero gravity laboratory, but the solar arrays need to rotate to follow the Sun. Since they are quite large, they are placed to the sides with rotating joints.This option mostly applies to fixed locations rather than vehicles. This is not a large enough distance to require high voltage lines for efficiency. Examples of future projects with longer power lines include mining water ice in shadowed craters at the Lunar poles. Solar arrays at the crater rim may have continuous sunlight while the mining area has none. So a transmission line can bridge the gap. Another example is a nuclear power source for nighttime power for a Mars Base. In that case the source is separated from the rest of the base for safety. Finally, large orbital habitats and industrial plants may use centralized generators that are lighter and more efficient, and need long power lines due to the size of the facility.

All types of wires need isolation from other system elements and each other, to prevent power leakage, shorts, arcing, and for safety. In a vacuum or non-conducting atmosphere, which the Earth's mostly is, isolation can be provided by mechanical gaps and spacing of wires. When the wires must be spaced close together, an **Electrical Insulator** can provide isolation, and combinations of spacing and insulation can also be used.

J. Electric Generator

An **Electric Generator** converts mechanical energy to electric energy. The two general types are a **Dynamo** which produces direct current, where electrons flow in one direction; and an **Alternator** which produces alternating current, where electrons flow in both directions in an alternating cycle. Most of the Earth's electric power is produced by large alternators. The mechanical energy enters the device via a rotating shaft, and an arrangement of magnetic fields induces a current in coils of wire. The mechanical energy can come from any of a number of sources. On Earth it is usually from high pressure steam or falling water acting on a turbine whose shaft is connected to the generator. In the case of steam, it is created by burning fossil fuels or a nuclear reactor, and more recently, from concentrated sunlight. A growing number of **Wind Turbines** are being used to produce electricity. The wind rotates aerodynamic blades mounted on a central shaft, and the shaft is connected to a generator

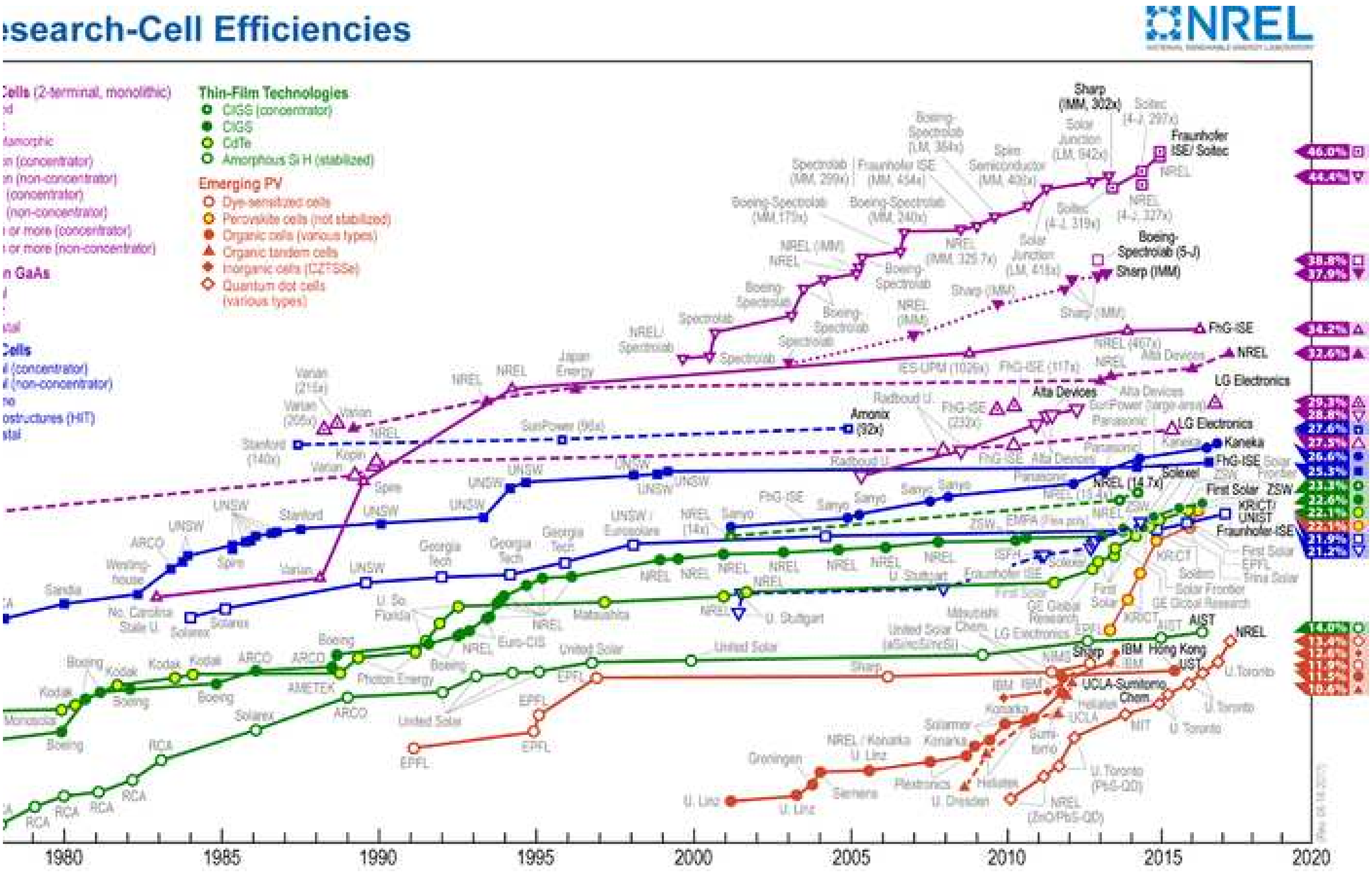
K. Magnetic Storage

Some space projects, such as electromagnetic launch from Earth, require very high electric power levels for short periods of time. These power levels can exceed what is available from the power grid. Magnetic storage accumulates energy over a longer period of time, then releases it quickly when needed. It uses a '*Superconducting* or high inductance/low resistance coil to store energy in a magnetic field. Superconductors eliminate resistance heating losses, but require cryogenic refrigeration to maintain the superconducting state. A large coil, cooled to lower resistance but not cryogenic, may be sufficient for some purposes. The energy, E (in Joules) stored in a magnetic field can be found fom

$$E = \frac{1}{2} L I^2$$

L is inductance in Henries, and I is current in Amperes. Storing energy in this way causes structural loads from the field back to the coil, so the total storage amount is limited by the strength of the structure. Some uses in space, like pulsed plasma propulsion systems, can benefit from smaller magnetic storage units to produce high power pulses fom a lower power steady souce.

L. Semiconductors



est research cell eficiency since 1975

Photovoltaic cells convert light, usually from the Sun, into direct current electricity using semiconductor materials. This technology is rapidly developing and has multiple materials and techniques (Figure 1.4-1). Conversion efficiency from sunlight of the best research cells, using multiple layers to capture different wavelengths, has reached 46.0% as of 2015. Production panels for use in space, made of multiple cells each, are near 30% efficiency, and the more common but less expensive single layer panels on Earth are typically 20% or less. Note that the efficiency on Earth vs in space are based on different solar intensity and spectra, because the Earth's atmosphee absorbs some wavelengths.

Thermophotovoltaic devices convert infrared and visible light from any hot object into electricity. They use similar semiconductors as photovoltaic cells, but optimized for the lower temperature source. A **Thermoelectric Generator** uses semiconductors to convert a temperatue difference into electricity. The most common use in space is generating power from radioisotope decay, in locations where solar panels are cumbersome, is not available all the time, or is too dim, such as beyond Jupiter. An isotope like Plutonium-238 produces decay heat on a steady basis, which thermoelectric cells convert to electricity. For space applications, pure efficiency is not the only significant measure. Variation with temperature, radiation exposure, and the specific power (W/kg) are also important. Given the trend of past improvements, it is expected semiconductor devices will continue to improve, at least in the short term. The latest data should be checked for current performance.

(The following are old references, and should be updated)

- Anonymous "Conference Record of the Nineteenth IEEE Photovoltaic Specialists Conference- 1987", New Orleans Louisiana, 4-8 May 1987.
- Anonymous "NASA Conference Publication 2475: Space Photovoltaic Research and Technology 1986: High Efficiency, Space Environment, and Array Technology", Cleveland, Ohio, 7-9 October 1986.
- Chubb, Donald L. "Combination Solar Photovoltaic Heat Engine Energy Converter", Journal of Propulsion and Power, v 3 no 4 pp 365-74, July-August 1987

M. Solar-Driven Turbine/Generator

In space, most electrical power so far comes from photovoltaics, since solar panels are lightweight and simple for small to medium amounts of power. For large scale power, **Brayton Cycle** turbines have been proposed, because of their potential high efficiency and low mass. The turbine shaft then drives an electric generator. The high and low temperatures for the cycle would be produced by solar concentrators and radiator panels. **Stirling** type engines have also been proposed for space use. Sunlight is abundant in space, and lightweight reflectors to concentrate it and feed a heat engine may be lower mass than photovoltaics.

(The following is an old reference, and should be updated)

- Spielberg, J. I. "A Solar Powered Outer Space Helium Heat Engine", Appl. Phys. Commun. vol 4 no 4 pp 279-84, 1984-1985.

N. Rectenna Array

A rectifying antenna, or **Rectenna** is an antenna that is used to convert electromagnetic energy into direct current electricity. A single antenna element can be a **Dipole**, with a **Diode** connected across the dipole arms. The incoming electromagnetic waves induce alternating currents in the dipole. Since diodes conduct in only one direction, a direct current is passed on. Many antenna elements are combined into an array to capture the whole of the incoming energy. Rectennas have been proposed as the receiving element of a long distance microwave power transmission system, such as from Earth orbit to the ground. Much more solar energy is available in space, which results in more net energy delivered on the ground, despite conversion losses. The beam can also travel from the ground to orbit to deliver power to a satellite.

The length of a dipole antenna scales with the **Wavelength** of the incoming energy. In principle, microscopic antenna arrays can be made by the same methods used for integrated circuits. This would allow for the direct conversion of infrared or visible light. Small scale antennas are in an early stage of research. Their advantage for long-range transmission is in a smaller transmitter for a given distance. Microwave technology, by comparison, is well developed, and high efficiency rectenna conversion has already been demonstrated.

Beam Sources

Entropy is a measure of disorder in a system. A directional beam of energy has low entropy because the waves are highly ordered (parallel). Useful work can be extracted from a low entropy system. This results in increased entropy (disorder), typically as random motion of atoms in thermal equilibrium and random thermal emission. Beams can be natural or artificial, and consist of electromagnetic waves or particles. Energy beams can be used for various kinds of **Beamed power propulsion**, or for powering more stationary activities.

O. Sunlight

At increasing distances from the **Sun**, around 14 million km or more, sunlight becomes highly directional. The source, which is 1.392 million km in diameter, then fills a small angular part of the sky. At the Earth's distance it appears 0.5 degrees in width. The small angle allows directional reflection as a controlled propulsion method. It also allows for concentration by lenses or mirrors, to generate high temperatures for industrial or propulsion purposes. This source includes direct use of sunlight, while the items under electrical sources are for sunlight converted to electricity.

The center of the Sun is at about 15.7 million degrees K, and has a core density of about 160,000 kg/m^3. Under these conditions Hydrogen undergoes **Nuclear Fusion** to Helium, releasing 3.846 x 10^{26} Watts of energy. This energy works its way from the core to the surface, where the temperature has fallen to 5,780K. At this point the intensity is 63.1 MW/m^2. At greater distances the same energy flow is spread over larger spherical surfaces, reaching 1362 W/m^2 at the Earth.

At a distance of more than 550 AU, the Sun acts as a **Gravitational Lens**, bringing the light of other stars to a focus. Light which passes farther from the Sun's edge is bent less, and comes to a focus farther away. This creates a radial focal line of concentrated light from every other star or light source in the sky. The same process happens around every other star that has visible neighboring stars. These **Star Lines** may be intense enough to be useful, since they concentrate light from the whole circumference of a star to a point.

P. Laser

A **Laser** emits light through a process of optical amplification by stimulated emission. The output of a laser is coherent and collimated, allowing it to be tightly focused, or travel long distances without spreading out. The output can also be in a very narrow range of wavelengths. Because of the narrow wavelength, it can be coupled efficiently to an absorber, or a high reflectivity reflector for that specific wavelength. It can also be coupled to a photovoltaic device with high efficiency. As an energy source for propulsion it can supply higher intensity light than natural sources like stars. High power lasers have been proposed for launch from Earth, but sufficiently high power ones to make that use practical do not yet exist. Lower power lasers can augment natural sunlight falling on spacecraft solar arrays. Very high power lasers focused by the Sun's gravity have been suggested to power interstellar vehicles, but that use woul far in the future.

Q. Microwave

This energy source involves direct use of the microwave beam, while item N. Rectenna Array converts it to electricity. A microwave beam can be absorbed and converted directly to heat, or it can be used to create photon pressure. Any suitable wavelength can be used to create a directional energy beam. However some wavelengths are absorbed by the Earth's or other atmospheres. Shorter wavelengths can be focused more easily, since that depends on the ratio of antenna size to wavelength. The efficiency of producing shorter wavelengths is typically lower, and generators with high enough power to be useful may not be available. Microwave band equipment is developed enough to not suffer from these limitations.

R. Neutral Particles

A **Particle Beam** is a collimated stream of high energy particles to deliver energy from one place to another. The concept was originally developed as a weapon, but less lethal amounts of energy can be used as a power source. Charged particles, such as protons in a **Particle Accelerator** repel each other, so a beam would spread out once it leaves the confinement of the accelerator. To prevent this, the charged particles are allowed to combine with electrons to form neutral atoms, or neutral particles like Neutrons are used. Particle beams are in an early state of development, and mostly for military use, rather than energy delivery.

Nuclear Sources

Nuclear energy sources involve a change in one or more types of **Atomic Nucleii**, with the release of net energy. The protons and neutrons in a nucleus are bound together by the strong nuclear force. A change in their arrangement involves typically a million times as much energy than rearranging electrons, which is what chemical reactions do. So nuclear energy sources are potentially very powerful. Although the Sun operates by nuclear fusion, we consider it a source of light energy. The fusion happens in the Sun's core, where it is not accessible, and what reaches us is blackbody radiation from the surface.

S. Radioactive decay

Radioactive Decay is the spontaneous change of unstable atomic nucleii by the emission of particles or electromagnetic energy. Unstable natural elements were created before the formation of the Solar System, most likely in supernova explosions. The less unstable ones, such as Uranium and Thorium, still survive after billions of years, and continue to decay at a steady rate. Artificial radioactive materials, such as Plutonium-238, are created in nuclear reactors or particle accelerators. They are more unstable, and thus decay faster (an 88 year half-life in the case of Pu-238). This element produces 500 Watts/kg of heat, when fresh, through radioactive decay, making it a useful energy source. It has been used for this purpose on a number of planetary exploration missions.Other elements with very long decay times in their natural state are too weak to use as energy sources.

[This is an old reference and should be updated]

- Lockwood, A.; Ewell, R.; Wood, C. "Advanced High Temperature Thermo-electrics for Space Power", Proceedings of the 16th Intersociety Energy Conversion Engineering Conference, v 2 pp 1985-1990, 1981.

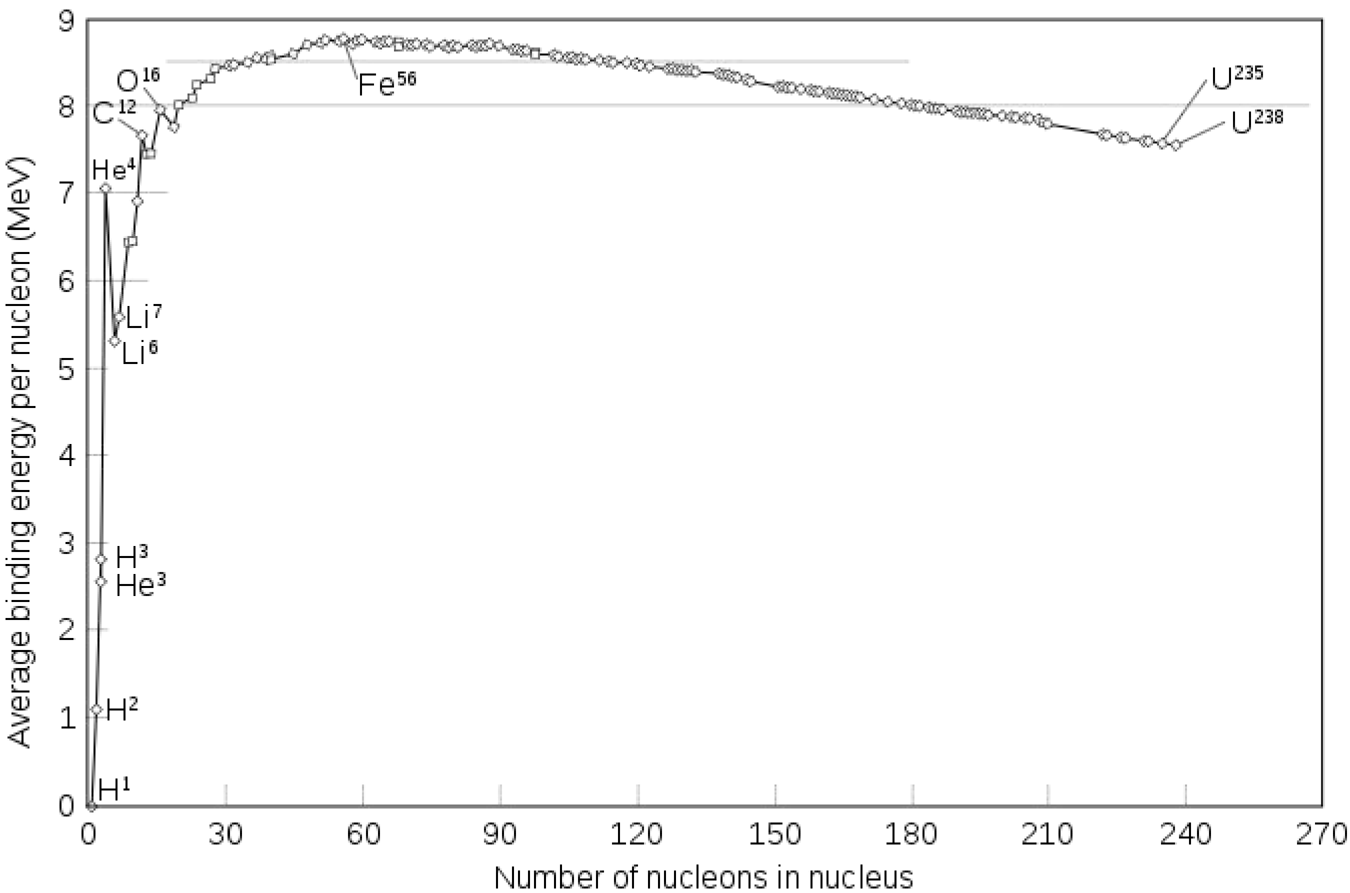


Figure 1.4-2 - Binding energy per nucleon for common isotopes, in MeV

T. Nuclear Fission

The low natural decay rate of some elements can be increased by artificial means. A **Nuclear Reactor** is a device for doing this in a controlled way for the production of energy. The two main ways to do this are **Nuclear Fission**, the splitting of heavy nucleii into smaller parts, and **Nuclear Fusion**, the merger of lighter nucleii to form a heavier one. The reason for the two types can be found in the **Binding Energy** per nuclear particle in a nucleus (Figure 1.4-2). A higher binding energy means the particles are more strongly held together and more stable, so energy can be released in forming that nucleus. The binding energy has a peak at Iron-56, so reactions from either the light end (fusion) or heavy end (fission) towards the middle both produce energy. Fission reactors are a significant source of electrical power on Earth. In space, a few small-scale reactors have been used, and work is in progress on developing larger scale versions with higher power demands.

[These are old references and should be updated]

- El Genk, M.S.; Hoover, M. D. "Space Nuclear Power Systems 1986: Proceedings of the Third Symposium", 1987.
- Sovie, Ronald J. "SP-100 Advanced Technology Program", NASA Technical Memorandum 89888, 1987.
- Bloomfield, Harvey S. "Small Space Reactor Power Systems for Unmanned Solar System Exploration Missions", NASA Technical Memorandum 100228, December 1987.
- Buden, D.; Trapp, T. J. "Space Nuclear Power Plant Technology Development Philosophy for a Ground Engineering Phase", Proceedings of the 20th Intersociety Energy Conversion Engineering Conference vol 1 pp 358-66, 1985.

U. Artificial Nuclear Fusion

Natural fusion occurs in stars, and the resulting light output has been addressed above under beamed power sources. This item is for artificial energy sources. Fusion has been achieved momentarily in nuclear bombs, but steady state operation has proved difficult. The most researched approach uses Tokamaks, which are doughnut shaped magnetic fields which contain a hot plasma. This approach has not yet produced a working device, although various research machines have been built or are under construction. A Tokamak type power reactor would be too massive for a reasonable propulsion system. For stationary projects it would be as reasonable on another planet as on Earth. A number of alternate intermittent and steady state fusion devices are under varying levels of research, but all at much lower funding than the work invested in the Tokamak type devices. Some of those might yield a lightweight enough device for propulsion.

All fusion reactions combine light atomic nuclei into heavier ones. As shown in Figure 1.4-2, the largest energy release is in the first few elements, from Hydrogen to Boron. What is required to achieve fusion is to bring the positively charged nuclei close enough together against their electric repulsion for the nuclear forces to take over. This requires the equivalent of millions of degrees K, or particle kinetic energy in the tens of kilo electron Volts (keV).

[These are old references and should be updated]

- Miley, G. H. et al "Advanced Fusion Power: A preliminary Assessment, final report 1986-1987". National Academy of Sciences report #AD-A185903, 1987.
- Eklund, P. M. "Quark-Catalyzed Fusion-Heated Rockets", AIAA paper number 82-1218 presented at AIAA/SAE/ASME 18th Joint Propulsion Conference, Cleveland, Ohio, 21-23 June 1982.

V. Muon-Catalyzed Fusion

Muon-Catalyzed Fusion is a method of catalyzing fusion reactions at temperatures far below the millions of degrees K otherwise required. A beam of muons is directed at a deuterium/tritium mixture, where they cause multiple fusion reactions. This heats the gas, which can then drive a generator for electricity. Although this method and more complicated systems based on it are sound from a physics standpoint, a practical system from an engineering point of view has not been developed yet. It must be considered a possible future energy source at this time.

W. Nuclear Explosions

Unfortunately, explosive **Nuclear Weapons** are all too well developed. Various concepts have been proposed to use their high energy output for space projects. These include a nuclear-powered launch device, where the explosive heats gas in an underground chamber. This then propels a projectile up a barrel. Another idea is to detonate small nuclear explosions behind a space vehicle, directly pushing it with the blast wave. These concepts are speculative at present, because there is no way to safely test them on Earth, and treaties prohibit nuclear weapons in space.

Matter Conversion Sources

In physics, **Mass-Energy Equivalence** is the idea that mass is related to energy by the formula $E = mc^2$. Since the speed of light, c, is a large number - 299.8 million m/s, the square is very large: 8.9875×10^{16} Joules/kg. This is equal to the output of a nuclear power plant for 2.85 years for each kg of mass converted to energy. In theory, total matter conversion provides the highest amount of energy per unit mass. In practice, however, this is not so easy

X. Antimatter

Antimatter is composed of antiparticles, which have the same mass as normal particles, but opposite charge and other properties. When a particle and anti-particle meet, they destroy each other and are converted to other particles or photons, releasing large amounts of energy in the process. Our universe is almost entirely made of matter. So use of this material as an energy source requires making it artificially. This requires at least as much energy as is later released by the annihilation. Antimatter is therefore an extreme type of energy storage. Antimatter is made in small amounts today in particle accelerators, and used for physics research. We do not have practical ways to make and store it in large enough amounts for space projects. Conceptually, a space vehicle would store some amount of antimatter, then use it to produce energy for propulsion. If the storage system is light enough, the energy per mass would then be higher than nuclear fusion or other methods.

[These are old references and should be updated]

- Hora, H.; Loeb, H. W "Efficient Production of Antihydrogen by Laser for Space Propulsion", Z. Flugwiss. Weltraumforsch., v. 10 no. 6 pp 393-400, November-December 1986.
- Forward, R. L., ed. "Mirror Matter Newsletter", self published, all volumes, contains extensive bibliography

Y. Black Hole

A **Black Hole** is a region of **Spacetime** with such strong gravity that nothing can travel from inside to outside it. Two forms of energy extraction are possible for black holes. The first is infall energy, generated as material in an accretion disk around the black hole heats by friction and emits energy. It is essentially converting potential energy into heat. Since the gravitational potential of a black hole is extreme, this can release a lot of energy. The second is **Hawking Radiation** by quantum tunneling from hypothetical quantum black holes. Black holes can form by the collapse of a large star at the end of it's life, or a sufficiently dense and massive region at the center of a galaxy. Quantum black holes are smaller, and hypothesized to have formed during the creation of the Universe. The nearest known stellar-mass black hole is 2800 light years from Earth, and quantum black holes have not been discovered, nor is there a known way to make them. So use of black holes for space projects is theoretical at present.

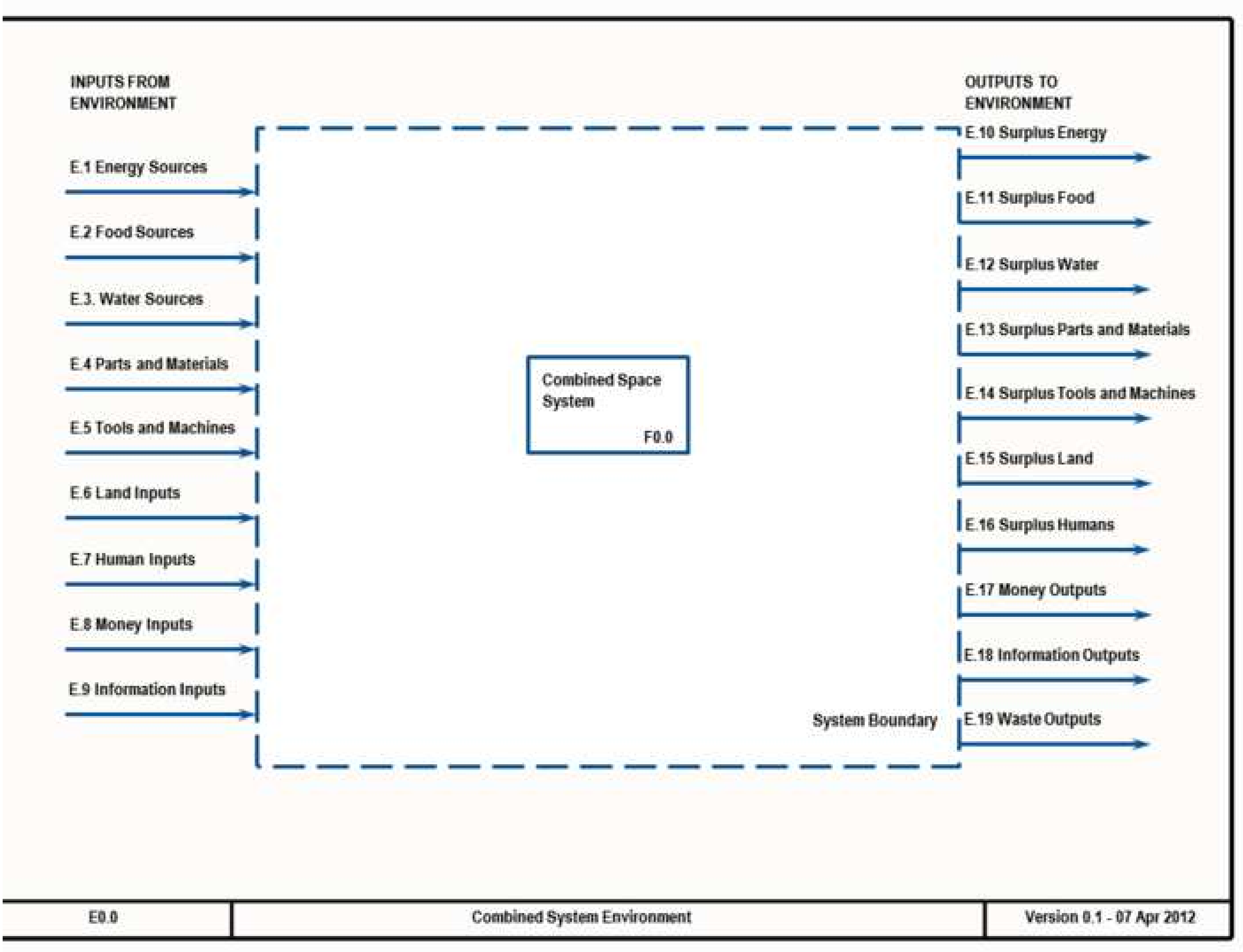
Section 1.5: Systems Engineering

Engineering applies scientific principles and other forms of knowledge to design, build, and operate systems which perform an intended function. It is a broad discipline, whose parts we will discuss later in **Section 1.7 - Engineering Specialties**. In a simple project, such as designing a bookcase for home use, a formal engineering process is not needed. One person can calculate the shelf loads and other parameters by themselves, with the help of a calculator and some reference data. Large and complicated projects, however, need the knowledge of multiple specialists, must satisfy multiple desired conditions and functions, and involve large amounts of time and money. A need then exists to coordinate the work, and ensure the final product meets the intended goal, in the most efficient way. **Systems Engineering** methods have been developed for this coordination task. They have become their own specialty field and are used in addition to the other engineering specialties. Systems Engineering can be used for any type of complex project. However, space systems are usually complicated enough to benefit from it, and systems thinking and methodologies are often used in this field.

Systems Engineering In General

What is a System?

Given an identified need or desire, how does one select the best design to satisfy it out of the infinite number of possible solutions? For a complex project, the concept of a **System** has proven useful. A System is defined as a functionally, physically, and/or behaviorally related group of regularly interacting or interdependent elements. They are distinguished from the rest of the Universe by a **System Boundary** (Figure 1.5-1). A system is not a physical entity, but rather a mental construct, created because of it's usefulness, by drawing a line or surface around a collection of elements. The elements have internal relationships to each other and form a comprehensible whole. The rest of the Universe outside the system is referred to as the **System Environment**, or simply the environment. Flows of many types enter and leave the system as **Inputs** from and **Outputs** to the environment by crossing the system boundary . The scope of a given engineering task is then defined by the system boundary, what crosses the boundary, and what is inside. Systems may contain smaller systems within them, which are called **Subsystems**. These may be nested to any level, but flows into and out of a subsystem must appear in the parent system, or at the top level in the environment. This rule may be called **Conservation of Flows** - that



ure 1.5-1: Example system diagram showing inputs and outputs to system boundary

flows do not appear from or vanish into nothing. Following that rule ensures that all the required inputs and outputs are accounted for.

Systems Engineering Method

A single person may have the time and knowledge to do a preliminary concept or design. A complete space project is usually too complex or would take too long for one person to do. So the **Systems Engineering** method can be used to help carry out such projects. Aerospace projects, including space systems, are particularly suitable due to their complexity, and were among the main ones for which the method was developed. The method focuses on how projects should be designed and managed over their entire **Life Cycle**, that being from initial concept to final disposal. Since it applies to the whole project, it is interdisciplinary, connecting tasks performed by Systems Engineering specialists to those of other engineering branches. Key parts of the process include:

- Breaking down a complicated project in such a way that the smallest pieces are simple enough for humans to design.

- Modeling the system so it can be analyzed and optimized, and comparing the actual physical system to the models.
- Control and track the information and design of the pieces and their relationships so the total system will do what you wanted.

Figure 1.5-2 illustrates in general the steps within the Systems Engineering process. The trend is from top to bottom, but we do not show arrows connecting the steps because it is not a strict linear flow. As results are obtained in any task, they can feed back to earlier steps in an iterative fashion, until a stable design solution is reached. So these tasks can happen in parallel, and applied across the different stages of the life cycle. The tasks can also be applied at different levels of detail. They are started at a general level. Once a stable configuration is reached at one level, it then is re-applied at lower levels until detailed design can be done on individual elements. At all levels, there is communication with design specialties, and with outside entities such as the customer, suppliers, and other scientific and engineering organizations. The steps are described in more detail in sections 3 and 4 below and on page 2. It should be noted that an organization capable of designing and building complex space systems is itself a complex system. While it is not often done, Systems Engineering methods can be applied to the organization itself to design and optimize how it functions, or to any complex system of any type, not just space hardware.

The systems engineering process is bounded by natural and human-made constraints. Many of the human constraints are not directly related to design in the way physical properties of materials are. These indirect constraints include economics, laws, and safety of life and property. The process is then also outward-looking, beyond the design itself. Other engineering specialties are more focused on the internal details of the design.

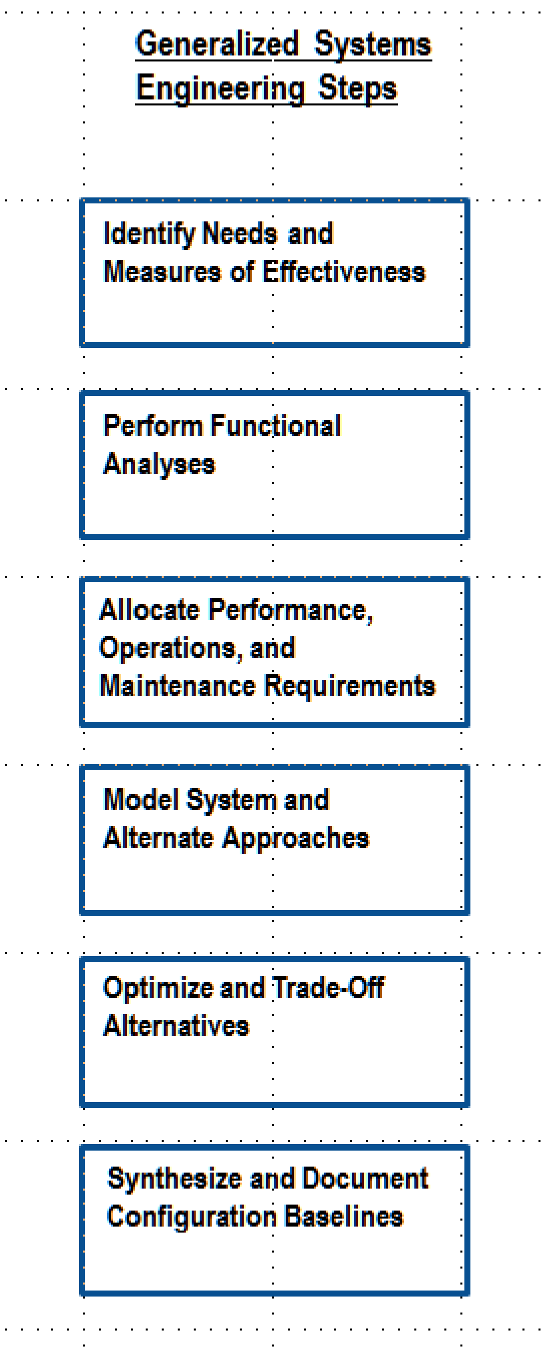


Figure 1.5-2

Large Scale Systems

It is not required that a single organization do all the Systems Engineering tasks. Some very large systems, that have been deliberately engineered, such as the US Interstate highway system, the Internet, and the US program to land humans on the Moon, involved many entities working together. A national system of government, human civilization, or the Earth's biosphere can be considered as very large systems in terms of having inputs and outputs, a system boundary, and an external environment. There is a growing understanding that such large entities are systems composed of many smaller systems, whether designed or not. Analyzing such large entities as systems can help with understanding how they function and determining if corrective action is needed. Although some attempts at designing governments have been made, they have yet to be done based on scientific and engineering principles. **Climate Engineering**, which is the concept of deliberately affecting the Earth's climate, is an example of biosphere level engineering projects. Doing them deliberately, as opposed to the inadvertent side effect of civilization, is still in the conceptual stage. More work has been done in the field of Economics in analyzing economic systems, and sometimes attempting to design or influence them.

Reference Sources

As a well-developed engineering specialty, there are a number of reference books, standards, and special methods and software used by systems engineers. They are used to understand and manage the interactions, and communicate the current state, of a complex project. The remainder of Part 1 of this book summarizes parts of the systems engineering method. This includes the elements of a system, engineering tools, involvement of other design specialties, and economics. A given program also has to be understood in the context of other existing and future programs. All of these tools and knowledge must be integrated properly for a new project.

For additional detail on Systems Engineering beyond what is in this book, see:

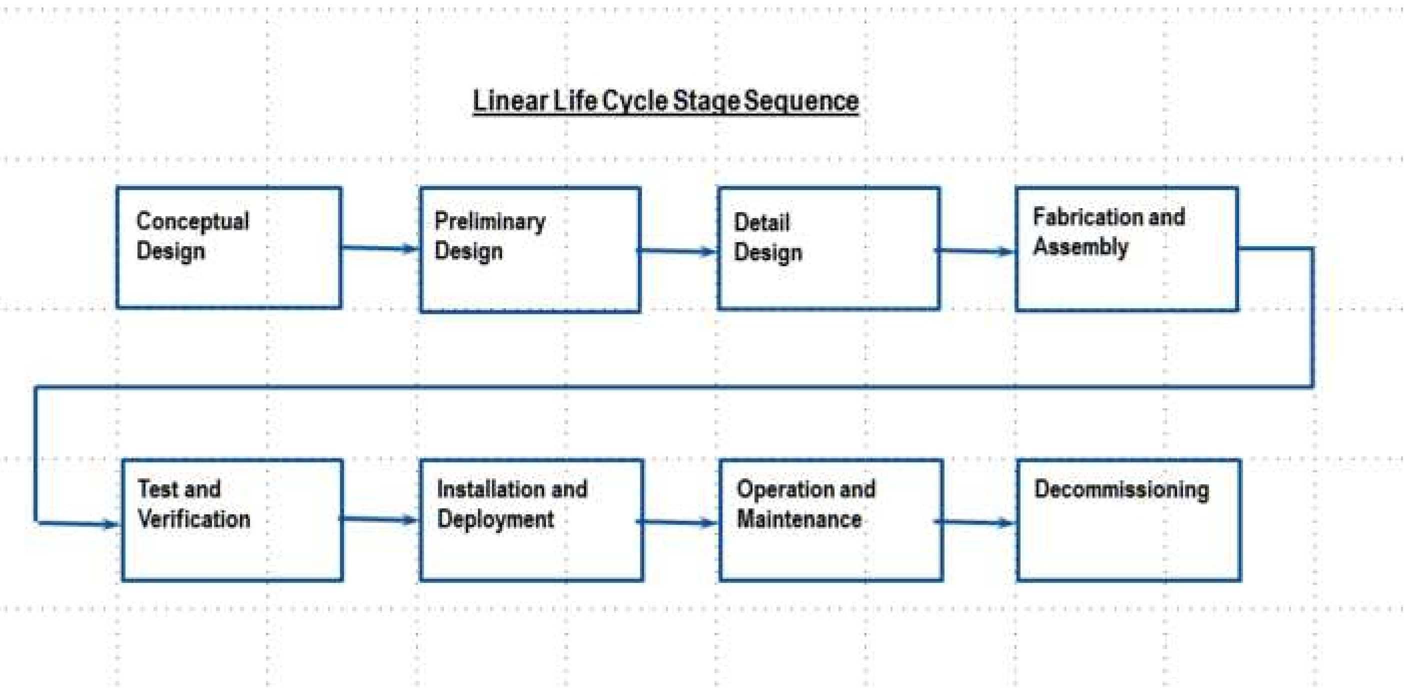
- BKCASE Project, **Systems Engineering Body of Knowledge** wiki, v1.8 - 27 Mar 2017.
- DAU Press, **Systems Engineering Fundamentals**, 2001.
- NASA, **Systems Engineering Handbook**, 2007.
- NASA, **Systems Engineering Class Materials** - Website developed since approx. 2008.

The System Life Cycle

Complex systems evolve through a **Life Cycle** much the way living things do, from conception to disposal. The life cycle is divided into a number of stages where different tasks are performed (Figure 1.5-3). The design stages (the first three boxes in the figure) can be organized in different ways depending on the nature of the system. These include linear, parallel, spiral, or closed loop sequences, or some mixture of these. The illustration shows a typical linear sequence. A spiral process repeats stages in increasing detail, while a closed loop repeats at the same level of detail. Beyond the design stages, the process is more typically linear from production, through test, installation, operation, and retirement.

Life cycle stages are used for two important reasons. First, the design process should consider all the later stages, so that the best total solution is found, rather than optimizing for just one part of a system's life. Second, breaking down a system by time is another way to simplify the design work, along with breaking it down by subsystems and components. The stages are further broken down into internal tasks which have inputs and outputs that connect them, and have decision points for when it is time to proceed to the next stage.

A life cycle is a time oriented view of an entire system. Other views of the same system include functional



ure 1.5-3 - Linear life cycle stages.

diagrams, which show what tasks it perfoms and their inputs and outputs, and a work breakdown, which tabulates the elements and sub-elements which make up the system. Which view of the system is used depends on the design task at hand, though all the views need to be kept current or the design process can become disjointed.

Life Cycle Example

The names, and the task contents, of a given project's stages can vary according to the needs of the project. However a somewhat standard linear flow is often used in aerospace engineering, including space-related projects. The stages and typical major tasks include:

Conceptual Design

- Identifying the need - what is it you want the system to do? This is embodied in goals and requirements.
- Establish selection criteria - how do you decide one design is better than another?
- Establish a system concept - this includes the main functions, operation, and maintenance of the system.
- Feasibility analysis - can the need be met at acceptable cost, schedule, and other parameters.

Preliminary Design

- Functional analysis - identify and break down the complex system into smaller functions and their relationships, including alternate arrangements
- Design Allocation - subdivide and assign requirements to lower tier functions
- Formulate alternatives - develop alternate solutions - what are the range of possible options?
- System Modeling - develop mathematical models of the system so variations can be assessed.
- Optimization and selection - making each option as good as it can be, then compare options and choose the best.

- Synthesis and definition - combining the selected options into a total design, and recording the configuration and requirements details

Detail Design

- Design - Once broken down to a low enough level, individual elements are assigned to engineering specialties or design teams to complete.Design includes physical hardware components and facilities, as well as software, operating procedures, training, and other non-physical elements.
- Integration - Design elements are combined into larger functional units that work together, up to the system as a whole.
- Engineering Models and Prototypes - Physical partial models and complete prototypes built to validate the design.

Fabrication and Assembly

- Production of components - For physical items, the step where you make the parts.
- Assembly - Putting parts together into complete elements.

Test and Verification

- Element and System Test - At each assemblylevel, testing that the assembly functions, then moving up to larger assemblies to the final product.
- Verification - Proving the system meets the stated design requirements, by a combination of test, demonstration, inspection, and analyses.

Installation and Deployment

After production, the system elements may need delivery, installation, and activation at the location they will be used.

Operation and Maintenance

- Operation - Using the system for the purpose it was designed, in the intended environment.
- Support - which includes operator training, performance monitoring, and logistic support.
- Maintenance - includes planned maintenance and unplanned repair, and in-place upgrades.

Decommissioning

When the system has reached the end of its useful life, the removal, recycling, and disposal of system elements, and return of former sites to their original conditions.

Life Cycle Engineering

As a process that applies across the whole life cycle, Systems Engineering is not just used in the initial design phase. Part of good design practice is to know when to stop designing. A design can always be improved with more work, but at some point additional work does not provide enough added improvement to justify it. At that point the design should stop, and the system progresses to the next stage, which is usually fabrication and assembly. With time, the original design assumptions for a project, such as the available technology level, or launch to orbit traffic levels, will change. The systems engineering process can then be re-applied to see if a design change, upgrade, or even complete replacement of the system is warranted. Even if the system was optimally designed when first created, future events may require changes. If the system was properly modeled and documented, then monitoring of these external changes will reveal when it is time to restart the engineering.

Requirements Analysis

Developing a new system starts with a desire or need which cannot be satisfied by existing systems. The needs and desires are expressed by a **Customer**, For systems engineering purposes, the direct customer is the person or entity who is paying for a project or can direct the engineering staff. For example, in the Boeing Company, that is the engineering managers and general managers of the company. The ultimate customers, which are airline passengers, cannot express their desires directly. So the company management serves as a proxy to express their desires as an input to the engineering process. Other methods, such as surveys, can be used to determine the desires of the ultimate customers.

The initial expression may be in the form of general verbal goals, system properties, levels of technical performance, and similar statements. The customer also will have some value preferences which describe what a better design is from their point of view. These can be things like "minimum cost", "minimum waste output", and "maximum efficiency".

The first major systems engineering step, **Requirements Analysis**, is the process of converting these general customer desires and preferences to specific measurable features which can be used for design and evaluation. Two main parts of this process are **Requirements Definition** and **Measures of Effectiveness**.

Requirements Definition

The highest level general desires are first converted to specific measurable features and values called **System Requirements**. These are later broken down into more detailed lower level requirements, which are assigned to logical elements of a system called **Functions** to perform. The assignment ensures that somewhere in the system all the top level goals are met. At the most detailed level a subset of the lower level requirements are assigned to a single function box. This now becomes the detailed design conditions for that function. Assuming the analysis has been carried to a low enough level, the detailed design of the element that performs the function can then be done with a reasonable effort.

The first step in requirements definition is documenting the original desire or need of the customer in as much detail as they can provide. We will use as an example the Apollo program to land humans on the Moon. That was expressed by President Kennedy as a **well known goal with a deadline**. That very general statement was not sufficient to design the hardware. The key task is to put all the requirements in forms that can be measured so that you can tell when you meet them. Experience shows if a desirable feature or parameter is not expressed and measured, it will not happen as desired. An example of this failure is the **Space Shuttle Program**. The original goal was to fly 60 times a year. Given a fleet of 4 Orbiters, each one had to fly 15 times a year, or one launch per 24 days. Subtracting 7 days on orbit and one day before and after flight for launch preparations and post-landing recovery, that leaves 15 days to complete ground processing. The stated goal was thus 160 hours for ground processing, composed of 2 shifts (16 hours) x 10 work days over two weeks, thus 14 days. This goal was expressed, but it may not have been included in the system requirements, and it definitely was not allocated to lower tier hardware and tracked at the lower levels like hardware weight was.

Only after the Shuttle was already flying was it noted that ground processing was taking too long, and efforts started to reduce it. At that point it was too late to make any fundamental changes in the design, and so ground processing never got below about 800 hours, about 5 times the original goal. This was a major contributor to the Shuttle never reaching its intended flight rate. In order to have reached their goal of 160 hours, processing time would have had to be allocated to sub-systems, such as landing gear or maneuvering thrusters, and then each subsystem designed to meet it's assigned time. Conversely to processing time, weight has always been a tracked parameter in aerospace systems, since airplanes cannot function if they are too heavy. The Space Shuttle had very detailed weight targets and a tracking system by component, with monthly reports. It more or less reached its design payload, which is the 1.5% of available launch weight remaining after the vehicle hardware and fuel are accounted for.

This example emphasizes why desired features must not only be stated quantitatively, but passed down for engineers to meet in detailed design, and tracked so you can tell if you are going to meet them. Measurable parameters can be a simple yes or no, for example "Does this airplane design meet FAA regulations?", or it can be a numerical value, range of values, table, formula, or graph indicating the range of acceptable values for that system characteristic.

Measures of Effectiveness

Desired features are often in opposition. For example, higher performance and reliability often come at the expense of higher cost. There are also alternate designs which have different amounts of each feature. Establishing **Measures of Effectiveness** is the quantitative method to account for these disparate features at the level of the whole system. Like requirements, they are derived from customer desires. In this case it is what features would be "better" when comparing one design over another. Since different features typically have different units of measurement, they need to be converted to a common measuring scale. This is done by formulas that convert each different value, such as cost or performance, to a score. These scores are given relative weights based on their importance to the customer. The weighted measures can then be used in a single mathematical model or formula to determine "better" as a total score, when the component values vary across different designs.

The value scale is often in a range such as 0 to 100%, or 1 to 10, but this is arbitrary. What is more important is a definite conversion from a measurable feature to a scoring value, and the relative weights and method of combining them to a total score. As an example, a value of 0% might be assigned to a payload of 15 tons, and 100% to a payload of 45 tons, with a linear scale in between, and payload given an importance of 30% in the total score. The total scoring system becomes a mathematical model of the customer's desires for the system. Getting a customer to define "better" in such detailed numerical form is often difficult, because it removes their freedom to choose what they personally prefer in a design in spite of the engineering solution. It is necessary, though, if you really want an optimal answer. At the least this process makes it obvious when the customer is over-riding the engineering process.

It should always be kept in mind that a particular design solution may not be "good enough" in terms of of it's measures when compared to existing systems. This can be found by including the existing system as one of the alternative designs being scored. In that case the proper answer is to stop development of the new system and stay with the existing ones. Often the cause is not enough performance improvement relative to cost, but other measures can result in a decision to stop.

Requirements Types

The following subheadings list major types of system requirements. Not all would be relevant to a given project, and others besides these might be important to the customer, so it is presented as a starting point for consideration. Each type can include more specific requirement values. The types listed here are linked and somewhat overlapping. For example high reliability and high safety generally go together. Requirements set limits on a design, and overlaps between requirements in effect overlap the range of limits they impose. This is acceptable as long as the designer understands the range of overlaps and interactions among the requirements. A particular design parameter will be governed by the strictest requirement when there is such overlap. An example from civil engineering is that earthquake, wind, and snow loads are all requirements to meet in a building design, and they overlap in that all of them affect the required strength of structural elements.

When broken down to lower levels of the system design, the requirement types and values will become more specific and detailed. Care is needed to maintain logical and numerical consistency across system levels. Requirements, or parts thereof, should not be inserted or dropped at lower levels. **Traceability** is the ability to follow the chain of requirements across the system levels, and is maintained by documenting how they are connected. This is necessary so you can prove satisfying the lowest level details actually meets the top level system goals.Historically the first two requirement types, performance and cost, were the primary ones considered. As systems have grown more complex and their outside interactions and side effects become better understood, the number of desirable features, and thus the number of requirements, have increased. This trend is expected to continue in thefuture.

Aside from the biblical **Ten Commandments**, requirements are rarely set in stone. Not all of them will be identified at the start of a project. As a result of interaction with the customer and feedback from the design process, they can end up modified. For example, a launch capacity of 10 launches at 100 tons each might be specified for a rocket, and later analysis show that 20 launches at 50 tons each yields lower total cost. The requirements would then be modified to reflect that. At any given time, however, the current set of requirements guides the engineering work. Over time, requirements become more firmly fixed, generally from the higher to more detailed levels in sequence. Changing a requirement forces rework of previous designs. So the cost of changing a requirement grows later in the process, and this tends to exceed any benefit from the change.

Performance

Performance requirements are measures of the primary intended function of a system. Every system must have at least one performance measure for what it does, and often there are a number of them. As an example, the design capacity for space transport systems is often expressed as a **Mission Model**. The mission model quantifies the system performance in terms of multiple parameters like dates, flight rate, payload dimensions and mass, mission duration, destination orbit, type of cargo, and maximum g-level. For a space habitat, performance might be measured in number of crew supported, levels of atmosphere, food supplies, and gravity, and total living volume. An industrial system might have requirements for **Throughput**, in tons per day of materials processed, and **Efficiency** in terms of (theoretical energy required)/(actual energy used). The particular performance measures which matter will vary by system.

An example mission model for the Apollo program might have started out as follows, with more detail added as the project progresses. Even in this early version, it lists a number of different performance measures that the design needs to meet:

Cargo characteristics:

Number of crew to the lunar surface: 2/mission
Maximum Stay time: 4 days/mission
Additional science equipment: 250 kg/flight
Lunar samples returned: 100 kg/flight

Mission Schedule:

First Flight: as early as possible but before Jan 1, 1970
Flight quantity: 10 to lunar surface (this was the original plan)
Flight rate: 4 flights/year

Performance requirements only address what a system does when it is operating as intended. It does not address what happens outside that context, such as

- In between active operation, such as the 80 days between Lunar missions in the mission model above.
- When the system fails, as did the Apollo 13 mission,
- Before and after the 5 years of manned missions.
- Interactions external to the program, such the supply of technical personnel for the project, environmental impact of the launches, or return of Lunar germs to Earth. The last turned out to be a needless worry, but the quarantine system for returning astronauts and moon rocks is not something covered in performance measures.

So performance alone does not encompass the entire system over the entire life cycle, and other requirement types are needed.

Cost

Cost represents the net resource inputs to a project from outside the system. Space projects don't use Dollars or Euro directly, but rather use them to pay for labor, materials, and services they do use. So cost is a measure of flows across the system boundary, rather than an internal property of the system. Every system will consume some resources during its life, but funding sources are not unlimited. So cost limits are almost always considered a requirement, whether implicitly or explicitly. Total cost over the entire life of the project is called **Life Cycle Cost**. This can be further broken down into development, production, and operations costs, and then accounted in much greater detail across the system elements. In addition to total cost, limits can be placed on spending rates. This is most explicit in government agency budgets, but even private projects have limits on spending per year. Some systems generate revenue to offset costs. When revenues exceed cost, the system as a whole generates a profit in financial terms. Revenue may be delayed until after the design and construction phases and the system begins to operate. The peak net cost accumulated until revenues exceed expenses is described as **Capital or Development Costs**. Customers generally want high performance and low cost, so the ratio of performance/cost is often a key measure for a project.

Compliance

Performance and cost requirements are set by the project's customer with the help the engineering team. **Compliance Requirements** are set by external human rules such as laws, regulations, codes, and standards. Human rules often set minimum requirements in areas like safety. This does not prevent a system from adopting stricter levels.Human rules are usually designed to prevent undesirable effects. For example, speed limits on driving are intended to reduce the frequency and severity of accidents. Compliance requirements exist whether or not they are explicitly incorporated into the engineering process. It is better to incorporate them explicitly to avoid later problems. Other requirements are set by nature, such as the minimum altitude for a stable Earth orbit. These do not fall under compliance, but are accounted for elsewhere. In the case of altitude, this might be a performance requirement that a rocket deliver the payload to a 250 km high orbit.

Technical Risk

Especially in the early stages of design, the engineering process may reveal gaps in knowledge, performance uncertainties, resources which are not available, or other issues which prevent selection, optimization, or synthesis of a design. These issues can prevent progress to the next stage of the project, or cause a final design which does not meet desired goals. Measures of these unknowns are given the general name **Technical Risks**. For example, a new technology which has not been demonstrated yet, i.e. a fusion rocket, would be rated as high risk, while a chemical rocket, which has decades of operating history, would be comparatively low risk. A mass budget considerably below past experience or with insufficient margin during preliminary design would be high risk. New research, modeling, or prototyping can be done to reduce the risks, or the system modified to avoid them. Before these risk reduction efforts the risks will still exist, and it is necessary to account for them. Otherwise you accept the alternate risk of the system not performing as desired or even at all.

Not every risk will be known at the start of a project, but sound engineering practice is to identify them as early as possible, and to allow for modifying development plans when they appear. Depending on how much new technology is included in a project, sufficient performance, time, and cost margins should be included for unexpected problems caused by technical risks. Technical risk is gradually retired during the design and production of a system.Once a system is operating, a small uncertainty remains for things like operating life or failure rates. These are not eliminated until the end of program operations. Even after disposal of a system, some environmental risk may remain. A prime example is nuclear waste, which is a hazard long after the reactor that created it has been demolished.

Safety

Safety is the state of being protected against adverse consequences to living things, or damage and destruction of inanimate objects. It is an inverse measure of other risks than those under the previous heading. So a higher safety level is measured by lower risks. Under the principle of "protecting the innocents", hazards to a crew that volunteers to accept a risk can be higher than those allowed to the public at large. A safe system, such as a nuclear power plant or passenger airplane, may have less than one expected accident during the system life. So safety often involves assessing low probability events. Requirements to maintain control of a system despite failures, inherent fail-safe design, design margins, backup systems, and redundancy can improve safety when properly implemented.

Reliability

Reliability is the probability the system will perform its intended function for a specified time period. The inverse is probability of failure. It is related to **Resilience**, which is the ability to function in the face of internal damage or external failures. It is also related to **Robustness**, which is the ability to function in the face of external or internal variables, such as line voltage or temperature. A closely related measure is **Availability**, which is the probability a system can start operating at a random requested time, or the percentage of a total time interval it can be operated. A high reliability system may require multiple units in place, so that at least the minimum required number are available at a given time. An example is passenger airplanes, which equire multiple engines for high eliability, in case one stops working.

Durability

Durability is how long a system can perform its intended function. It is typically measured in terms of service life. Service life of components, and of the system as a whole, is related to their maintenance. If not enough maintenance is performed, the life is reduced. Once the life of the item has been reached, it will need repair or replacement. Service life can be measured in terms of number of uses, operating hours, or calendar time. Durability is related to the economics concept of a **Durable Good**, one that yields utility over time rather than being consumed in one use. A passenger airplane has high durability, because it can operate for decades and tens of thousands of flights. The opposite of durability is consumption. The fuel used in the airplane is consumed (used only once), and is therefore not durable.

Quality

Quality is a measure of how well a system meets expectations. One aspect of quality is a measure of the lack of variability or defects in the design and manufacturing stages of the life cycle. Variability and defects increase the chance that performance will fall outside required levels. Another way to put it is conformance to initial specifications. Wear or defects caused during normal operation are not a quality problem unless they are unexpectedly large. Normal wear would fall under maintenance requirements. Another quality factor is parameters like signal-to-noise ratio and error rates in electronic devices. Noise and quantum effects are natural variations which cannot be eliminated, but a large margin above those variations in effect reduces variability and increases quality.

Sustainability

Sustainability is the capacity of a system to endure. For example, does a system consume a scarce resource or generate a waste output that limits it's long term use? If so, it is not sustainable over the long term, although it may last a desired system life. A current example is hydrocarbon fueled rockets. If obtained from fossil sources, they are both limited in supply, and cause unwanted change to the atmosphere. If they are produced as a biofuel, they can be sustainable.

Community

These are requirements to have a positive, or at least minimally negative, impact on the surrounding human community. This might mean employing local staff, or avoiding traffic problems during a shift change, or noise impact from rocket launches. Positive educational impact is another community effect.

Environment

Like community, these requirements relate to impact on the surroundings, but in this case the non-human portion. For space projects a key environmental requirement is to avoid contamination, either biological, chemical, or from radiation, both forward and backward from the ends of the mission.

Manufacturing

This type of requirement covers items such as how many sources are there for a given manufacturing method, or how exacting the tolerances are. In sum they measure how easy or hard the system is to produce.

Testing

These requirements deal with the types and quantity of tests required for a system. Development tests occur during the design stage, qualification tests occur during approval, and periodic inspection and testing may be needed during the operations stage.

Maintenance

These requirements include parameters like hours and costs to maintain the system in an operating state, probability of system failure, and levels of spares required. A system can fail without safety risk. For example, your car may not start, which is different than the brakes failing to operate. The more often that items fail to operate, the more spares are needed in stock, and the more time and money to repair or replace items. So the various maintenance requirements are linked. Maintenance requirements can be divided into preventive, which is before something stops working, and corrective, which is after.

Flexibility/Adaptability

This is the ability of the system to adapt to new tasks or functions, or different performance levels for the original tasks, than first designed for. Similar requirements come under names like **Extension**, which is adding new tasks or functions while keeping the original ones, or **Agility**, which is concerned with how fast the system can adapt. **Reconfiguration** is the ability to change the arrangement of a system to perform a different function. An example is the Curiosity rover, which had one physical shape and software load to cover travel to and landing on Mars, and a different arrangement of wheels, camera mast, and software for surface operations. The change of state was accomplished by a combination of mechanical design, planned sequence, and new software upload.

Scalability

This is the ability of the system to change in size either by scaling the size of a unit, or by installing more units. There is always a limit to scaling imposed by some physical constraint. If the system can be scaled to meet the full demand for it without reaching scaling limits, it can be said to be scalable. **Modularity** is a related parameter, concerned with how separate the elements of the system are, and how easy it is to replace them with other elements of the same or different types. Instances of modularity can be horizontal - at the same level of a system, or vertical, as in the layers of the Internet Protocol stack.

Evolution

This requirement type considers how well the system can change over time to a different type of system. It is related to flexibility, which is more about changes while staying the same kind of system. **Redesign** is concerned with the difficulty and cost of making changes.

Usability

Usability requirements deal with the interface between humans and system elements. When a system can be used without great amounts of planning, preparation, physical strain, or training it is said to have high usability.

Interoperation

This is a measure of how the system fits with other systems. For example, a new airplane with doors that did not fit existing gates, or a computer network that exclusively uses a new protocol that nobody else uses would fail in this parameter. **Compatibility** is more concerned with the direct interfaces between systems, such as the output of a computer video card matching the input to a monitor. These features are more prominent in the information technology fields because of the sheer number and variety of hardware and software elements which must work together (with varying degrees of success).

Openness

This is the degree to which a system is composed of proprietary or secret elements compared to open, public, or standard elements.

Continued on page 2 →

Section 1.5: Systems Engineering (page 2)

← Back to Page 1

Functional Analysis

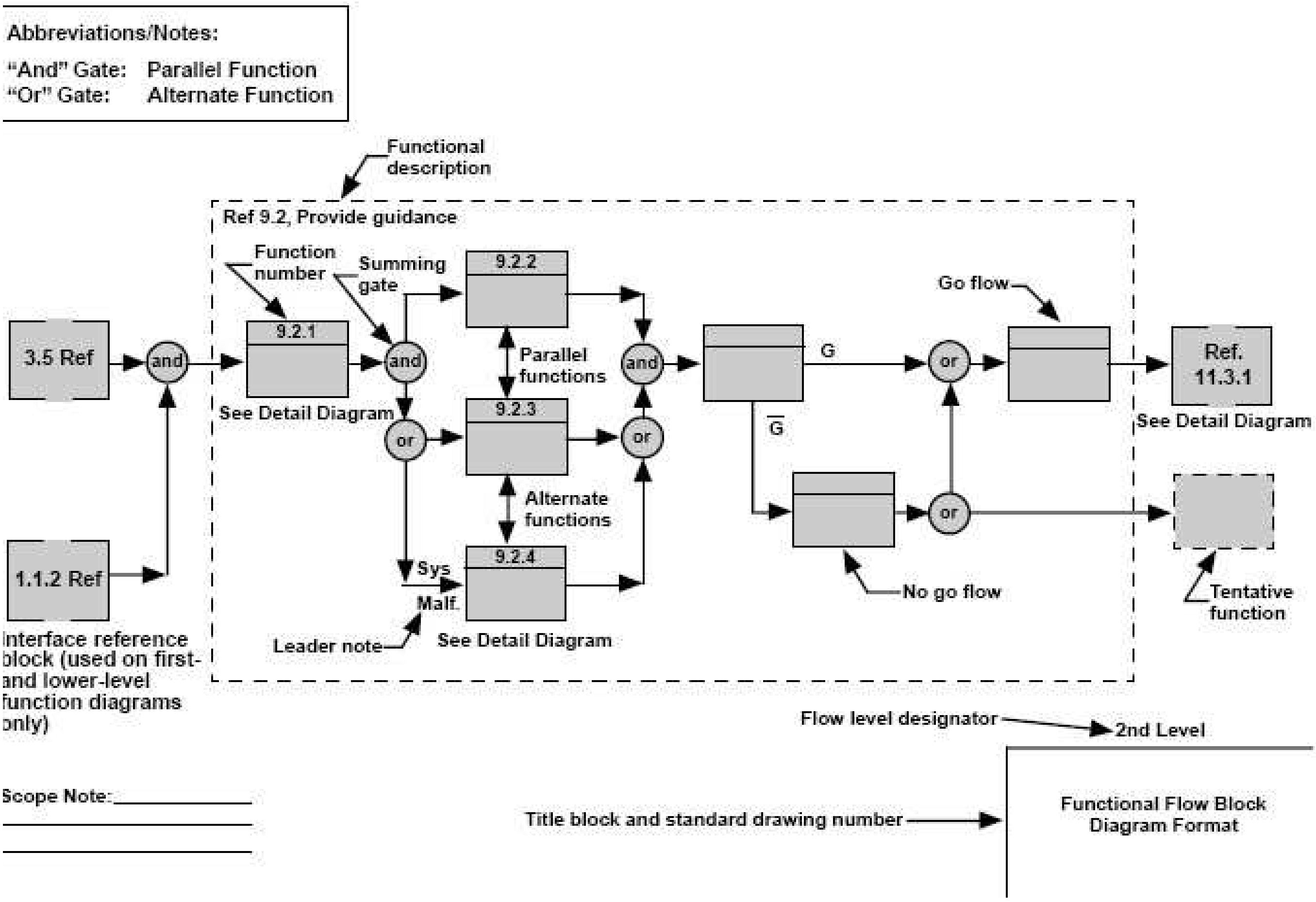
Once requirements have been developed to a certain level (see page 1), the next step is **Functional Analysis**. Engineering analysis in general is the breaking down of an object, system, problem or issue into its basic elements, to get at its essential features and their relationships to each other, and to external elements. Analysis includes developing abstract models or performing calculations for the component elements of a system, to help arrive at a complete and optimized design. Functional analysis is a breakdown on the basis of **what** a system does, in terms of the functions it performs or a sequence of operations. This is before considering **how** it is performed. "How" is a design solution, which we don't want to choose prematurely. Instead we want to consider all the alternatives and optimizations, which we do in later steps of the process. Prior to selecting the best design, there may be multiple functional breakdowns, at least one for each system concept, and often alternate versions within a concept. The details of these steps and their interactions are recorded in the form of diagrams and models, which can then be used for calculations and assessments.

Concept Identification

Although we don't want to prematurely select a design solution, we do have to generate alternate concepts for how the system will function. At the level of the system as a whole it is difficult to define requirements and measures without an idea of how the system will work. Concept identification involves identifying alternate approaches to how it will operate, and synthesizing one or more system concepts for each potential approach. System level concepts give the general approach to design and function, without specifying exact values of parameters or what components will be used. For the Apollo program, for example, there was a choice of "Direct to the Moon" vs "Lunar Orbit Rendezvous" as mission concepts, with the latter being the one actually chosen. System concepts may include major variables such as type of propulsion (i.e. chemical or nuclear), service life (one or multiple missions), and supply concept (i.e. closed or open loop life support). At the least it covers what main tasks the system will perform, and how it will be operated and maintained. Once the system concepts are established, the process of analysis, optimization, and selection can begin to find the best version of each competing concept, and then compare and select among the concepts to carry to the next stage of development.

At lower levels of the design process, this step is repeated when there are multiple possible approaches. An example would be thermal protection for re-entry. An ablative heat shield burns off some material each time, so checking thickness and periodic replacement would be part of the necessary operations. A metallic heat shield might not burn off, but suffer cracking from heating and cooling cycles, and require different types of inspection. So when listing design alternatives, it is not just ablative vs metallic that is important, but also how that choice affects the total flow of operations in the system.

Functional Diagrams



..5-4: Complex Functional Flow Diagram.

Functional Diagrams are prepared during analysis to illustrate the component operations a system perfoms and their inputs and outputs. They are a model of how a system operates in visual form. They start at the top level with external inputs and outputs which cross the system boundary, then are broken down into multiple levels of detail. They are often in a step by step time sequence, but can be more complex networks of operations with decision points and loops (Figure 1.5-4). For example, for an airplane, the main functions would be Load Passengers and Cargo, Taxi to Runway, Takeoff, Fly, Land, Taxi to Gate, and Unload Passengers and Cargo. Each of these main functions are further divided into smaller steps, then assigned to system elements to carry out. For example, the landing gear might be assigned multiple functions such as "absorb landing loads" and "provide steering for taxiing". Those then become requirements for detailed design and testing of that element.

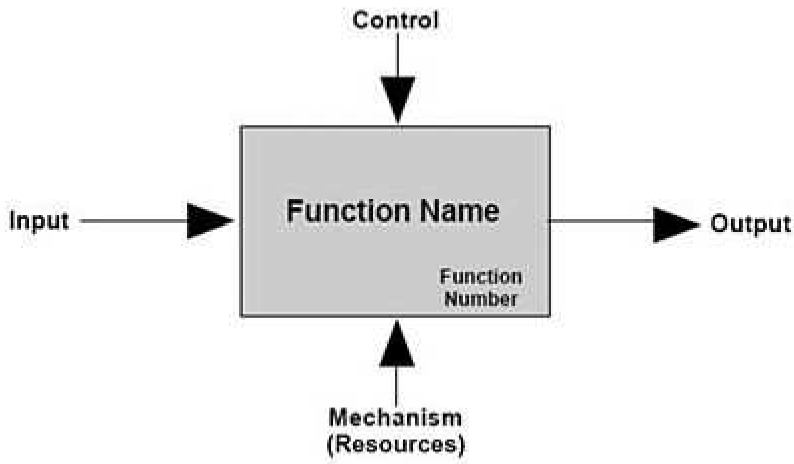


Figure 1.5-5: Single Functional Diagram Box

Individual functions in a diagram transform inputs into outputs (Figure 1.5-5). The diagram typically shows functions as boxes and input/output flows as arrows connecting the boxes running from left to right. Flows can contain any sort of item, including information, matter, energy, labor, etc, or a combination thereof. They may divide and combine on a diagram, but the divided flow must sum to the contents of the undivided one. This derives from the physics concept of conservation laws, where matter and energy do not arise out of nothing. Similarly the flows within a system do not arise from or vanish into nothing, they must enter from outside or be converted by a functional task. By following the **Conservation of Flows** logic, then all the inputs and outputs of a system will be accounted for

Control inputs regulate the operation of a function. By convention they are shown entering the top of a function box. Mechanisms perform the function, but are not transformed themselves, and are shown entering from the bottom. A mechanism example is a stamping press, which converts flat steel blanks to shaped stampings. Tne blanks and stampings are the inputs and outputs, respectively. For a complex system, the diagrams form a hierarchy, with one box on a given level expanded to a full diagram with multiple boxes at the next level down. Developing the levels of diagrams is a continuing task done incrementally, rather than all at once. The diagrams are a way to record and communicate the structure and operation of a system. They allow numerical calculations, for example noting the time required for each step to find the total operation time, or summing staff required for each function to get total staff needed to operate the system. Functional diagrams can also be converted to mathematical simulations of system operations, typically with computer software made for that purpose. Any amount of description or other information may be attached to items in a diagram, by means of a unique function or flow reference number. By convention, expanded lower level diagrams use the same number as the parent box (i.e. 9.2), with another period followed by another number (9.2.1, 9.2.2, etc.). This is not required, but it makes tracing the connections between diagrams easier

Requirements Allocation

The third major step in the systems engineering process is **Requirements Allocation**. To ensure all the top level requirements will be met, they are assigned to one or more functions to implement. The assignment may be the whole requirement, or by dividing it into parts and then assigning the parts to separate functions. Allocated requirements are documented in lower level requirements documents or specifications that apply to parts of a project. **Traceability** is being able to trace the links between lower and higher level requirements and the logic of how they were generated. At the lowest level, a subset of the requirements are assigned to a particular hardware or software item, skilled staff, procedures, facilities, interfaces, services, or other elements of the final system. With a complex project, software tools become very useful for the requirements allocation and tracking process. They can help manage the mass of details, and ensure everyone on a project has the most current information.

Requirements allocation is not a one-time task, although it is weighted towards the early stages of a project. As design and testing make progress, they can provide feedback and adjustment to the assigned requirements. These changes propagate to higher levels, and by tracing back their impacts, you can determine how they affect the top-level goals of the project. Changes can also have sideways effects at the same level. For example, an increase in weight of one part of a system may require weight-saving efforts elsewhere to not affect top level performance.

System Modeling

The next major step is to model the system and alternate approaches to the design. Various methods are used to model the design and configuration of the system elements. Traditional ones include two dimensional diagrams (blueprints) and physical scale models. These methods can help visualize a system, but are not easy to modify, derive parameters, or perform simulations. The trend is towards integrated software modeling, where software tools model and simulate multiple aspects of the system, or communicate from one tool to another. In software tools, a system is represented as data and mathematical relationships, which makes it much easier to change, optimize, and evaluate.

Input/Output Model

Input-Output Models were first developed for quantitative understanding of the total flows in an economy. They can be applied to any system, not just economic ones, for determining if all the inputs and outputs of a system add up. It can be visualized as a spreadsheet with the elements of a system as rows, and additional rows for items outside the system. Types of inputs and outputs are in columns. Each component requires inputs such as power, data, fuel, etc. It also produces some kind of output. The purpose of a model is to see if your system as a whole has closure and balance. In other words, are all the inputs matched by outputs? Are there missing components identified by missing inputs? Is the size or quantity of a particular element in the system the correct size/productivity? Will the system as a whole produce the desired output, and if so how much? These are really all questions of accounting. Rather than counting everything in money, this type of spreadsheet does the accounting of each type of input/output/resource/supply separately as categories. Note that human labor is one of the input types.

A model, such as an Input/Output Model lets you actively see how a change in any one component, such as a new design, impacts the system as a whole. By summing the flows of the component functions in a model spreadsheet (or other computer model) you can immediately find changes to the rest of the system components and the totals for the entire system. The Input/Output model and functional diagrams both model aspects of the same system, and may be combined within a single software tool or database if it can represent all the details of a system in sufficient detail. This is desirable for plotting how changes in system components affect the system as a whole.

Functional diagrams at a basic level are maintained as static drawings, and input/output models can be an actual spreadsheet. Using the same numbering system and structure for the diagrams and models maintains the relationship between them. They are both representing the same system, just different aspects.

Optimization and Trade Studies

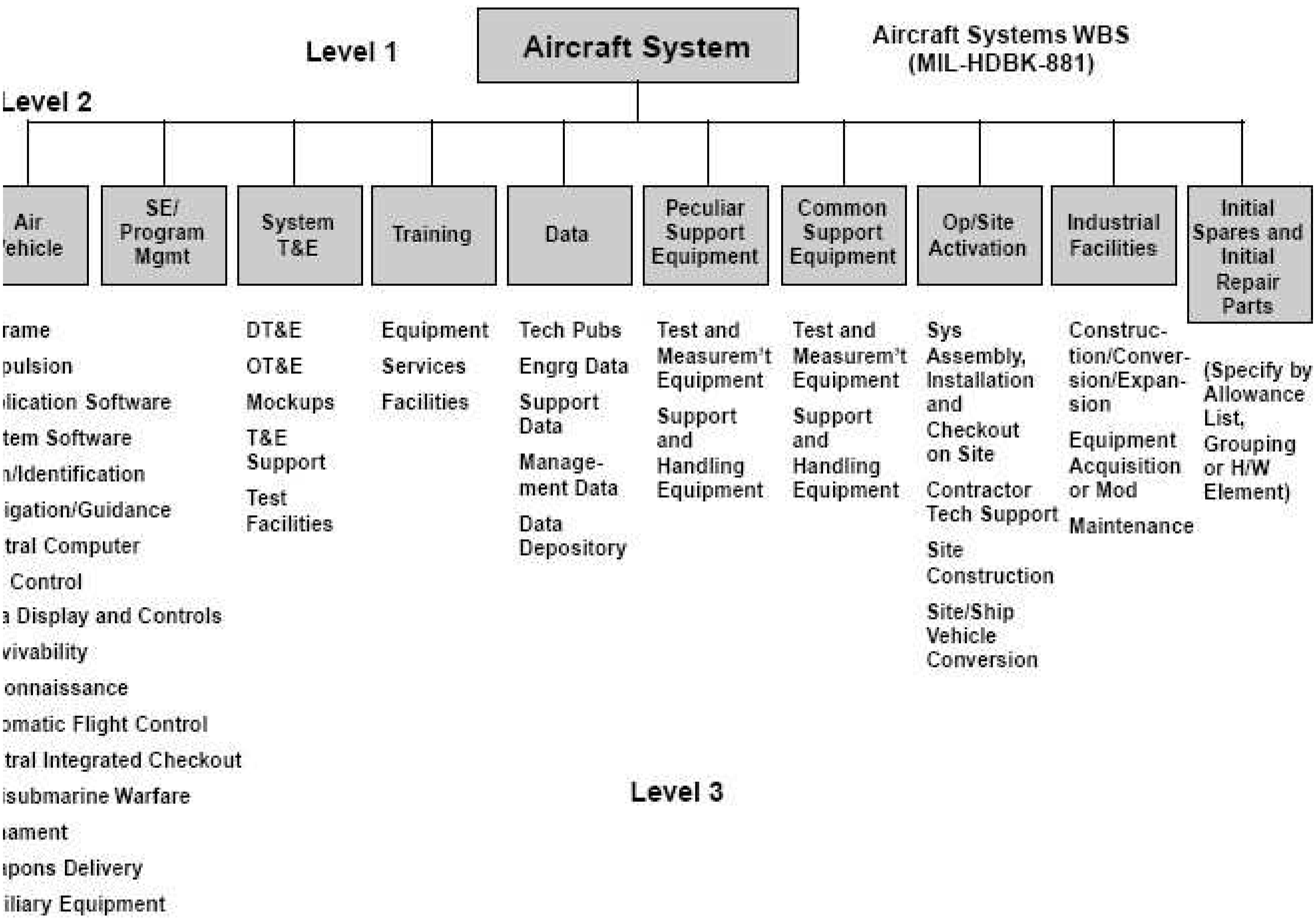
Optimization and selection is done at all levels of engineering design. In the Systems Engineering process it is first applied at a high level to concepts before detailed design is performed. Optimization is varying parameters within a single concept or element in order to find the best values for those parameters. **Trade Studies** compare different concepts in order to select the best one. Different parameters like weight, cost, and risk cannot be directly compared. So they are scored by measures of effectiveness (see page 1). The concept or optimized parameters that yield the highest score is the "best option". In the early stages of design, there will be larger uncertainties in parameters like weight and performance. Finding how much effect variations in such parameters have is called **Sensitivity Analysis**. Knowing those will guide which areas to work on to reduce uncertainties.

If the difference in evaluation score between two concepts is sufficiently more than their uncertainties, the lower scoring one may safely be discarded. If the scores are within the range of uncertainties, both should be worked on in more detail until a clear winner emerges. If the effort to reduce uncertainty is judged more than the uncertainty reduction is worth, then one of the competing choices can be selected arbitrarily. Note that optimization of a system as a whole may not mean optimization of each individual part, since the parts can interact in a complex way. Once the optimization and selection is completed, the results are recorded and used to update the system concept and current design configuration.

Synthesis and Documentation

The last major step in the systems engineering design cycle is synthesis and documentation. In systems engineering, "System Synthesis" is assembling the results of completed analysis and studies into a coherent design. The design for a complex system typically includes multiple items of hardware, software, facilities, etc. Each separate item is referred to as a **Configuration Item**, and the current state of that item's design at any given time is called a **Configuration**. **Configuration Management** is the task of documenting the current state of the design and analysis work. This is necessary to coordinate the work for a complex design with many people involved. Otherwise some work would be based on obsolete data or incorrect assumptions. Other documents included in recording the work are Requirements, Specifications, Study Reports, Simulation code and results, 3D Models, and any other data and notes created in the course of the work. All of this is kept as a base for further work, if later changes are needed, or if questions or problems come up. Design data is also needed for later project stages like production.

Work Breakdown Structure (WBS)



..5-6: Example WBS Drawing.

A common method to document a system is to index all the requirements, plans, drawings, analyses, reports, budgets, work logs, and other data by a numbering system called a **Work Breakdown Structure**, which covers all the elements of the system across it's life cycle. In modern projects the actual data is mostly stored electronically, but a WBS helps organize and find particular items in the same way classification systems for books are used to organize libraries.

A WBS is a hierarchical table or drawing showing all the parts of a complex system and their successive division into smaller parts until you reach the level that specialty engineers can do the detailed design. It gives structure to what would otherwise be an amorphous mass of design work. The WBS serves as a tracking method and index, so that people working on different parts of the project tell they are talking about the same items. It also serves as a method to collect and file the engineering data as it accumulates, assign tasks to individuals and groups, and track progress and costs. The WBS is often derived from the functional analysis of the system.

In theory a WBS can be structured in any way you choose, but usually each level of division within the structure has a common basis.Examples include:

- By location on Earth or in Space
- By type of function, such as Production, Operation, and Transportation
- By type of element, such as Data, Software, Hardware, Facilities, and Staff
- By end item product, such as a launch vehicle or lunar base
- By Subsystem, such as structures, mechanical, or electrical
- By time sequence, such as Phase I, or Version 2 Upgrade
- By type of data such as drawings, analyses, or reports

The basis to use depends on what makes sense for the project, but a consistent structure, such as all second level divisions are by end item, makes the overall structure easier to understand and use. It is more important that everyone on the project use the same structure than exactly how it is divided up. Maintaining the structure is often assigned to systems engineering specialists because it is related to the other tasks they perform. Each part of the structure is given a number or identifying key, typically using decimal points to distinguish levels, i.e 1,2,3, ... for the top level, then 1.1, 1.2, 1.3, ... for the parts that make up the next level below item 1, and so on. This is not the only way to do such structuring, but it is commonly used and easy to understand. The following section illustrates some of the ways to arrange a given WBS level. It is not an exhaustive list.

WBS by Item Type

This example is for an automated factory that consists of operating data, software, and hardware, facilities, and staff:

1.0 Operating Data Components

- 1.1 Design Standards
- 1.2 Manuals and procedures

2.0 Software Components

- 2.1 Design Software - A great deal of design software already exists. The specific need for advanced manufacturing is to design parts so they can fit the production capability of given machines, and supply processing and assembly instructions for individual and collections of parts. This may required modifying or adding to existing software.
- 2.2 Work Order Software - Takes an incoming product design in the form of CAD files, compares it to the factory capabilities and inventory and generates a work order list of tasks for each machine, parts and materials to order, etc. Work orders are then scheduled among the various components.
- 2.3 Machine Driver Software - Each type of automated machine requires specific driver software to control how it operates, and to collect data back to track progress and other purposes.

3.0 Hardware Components

- 3.1 Storage - Materials, parts, and assemblies need to be stored when not actively being worked on.
- 3.2 Materials Handling - To transport items from one location to another
- 3.3 Production Machines - Turn raw materials into inventory stock or finished parts, possibly using several machines for different steps.
- 3.4 Assembly Machines - Convert a collection of parts into a finished item. This will generally involve one or more robots.
- 3.5 Inspection and Observation Hardware - To test items and oversee operations.

4.0 Facility Components - This includes modification of the surroundings, controlling the factory environment using buildings, and supplying utilities, but not specifically producing any items.

5.0 Staff Components - Humans are not components to be designed, but rather selected and trained for required skills, and then supplied in needed numbers.

WBS by Subsystem Type

This example is a typical set of subsystems for space hardware, and also lines up with design specialties:

- **Structures**
- **Mechanical**
- **Power and Electrical**
- **Propulsion**
- **Thermal**
- **Data**
- **Communications**
- **Sensors**
- **Displays and Controls**
- **Internal Environment**
- **External Environment**
- **Crew Support**
- **Maintenance and Repair**

Section 1.6 - Engineering Tools

Space system engineers use a wide variety of tools to do their work. The most important of these is their own knowledge and experience, which we hope this book helps upgrade. The work includes producing designs for a project; then recording the designs and other information in formats that can be shared with other people, or for use by computers and machines as input. For simple or early stage work, some reference books, a scientific calculator, and a pad of graph paper may be all that is needed. For the bulk of the work as done today they typically require a variety of data sources, computer workstations and more powerful supercomputers or networks, and specialized software. When the work gets past design into R&D, prototyping, manufacturing, and test, they often use physical tools and specialized test equipment to measure performance and collect data.

Engineering Data

No engineering design can be done without input data of some form. It can be determined internally, but more usually obtained from outside sources. Types of data include:

- **Engineering Codes and Standards** - These are documents that specify required or accepted methods and features for a design. For example, **Building Codes** embody accumulated experience in how to design and build safe and sound buildings. Adoption of building codes by governments gives them the force of law - they must be followed. **Technical Standards** are formal documents that establish uniform and accepted engineering criteria to be followed. For example **ASTM Standards** for the composition and strength of steel do not have the force of law of themselves, but allow steel suppliers and engineers to work together because both know what is expected from a given alloy grade. Standards may be incorporated by reference in laws, regulations, or contracts. Large engineering organizations may develop their own internal design standards based on their experience, so that more consistent results are obtained and new staff can be trained.

- **Handbooks, Textbooks, Monographs, and Journals** - Handbooks are compilations of useful information in a particular engineering field. They are often written by multiple contributors and updated periodically. An example is the **Handbook of Space Technology**. For students, handbooks are often quite expensive, so it is suggested to find them used or in a library collection. Textbooks are intended to teach a subject, like this Wikibook. Monograph means "one writer", books written by one or a few authors. They are typically about advanced topics, primary research, or original scholarship. Journals are periodical publications containing shorter articles than monographs, reporting new research or reviewing the state of the art. There are a vast number of books and journals covering every engineering topic, so it is impractical for an individual to collect them unless for a very narrow specialty. A good technical library can provide access to all these works.

- **Supplier Data** - One of the basic rules of efficient design is to not repeat a design if someone else already has. Many designs will incorporate parts or subsystems that already exist and are made by someone else. The suppliers of such items have literature and documentation about what they supply, and often will consult about the use of their products.

- **Online Data Sources** - A huge amount of data is online these days, but the quality is variable. Since incorrect data can lead to space systems failing, these data sources should be carefully selected for quality. Online data also changes quickly, so any links we give would soon be out of date. A good approach is to use a search engine, and understand how to define the search terms to get a specific result.

Computer Hardware

Historically engineers worked at large drawing tables or desks where they could produce the drawings and documents that represented a design. Such methods have been largely replaced by computer workstations for several reasons. Computer systems can communicate changes much faster than paper-based methods. They can represent designs in three dimensions, which was difficult on two-dimensional paper. And finally, computers can perform analysis and simulation of a design vastly better than hand methods. At one point mainframes and engineering workstations were specialized and expensive equipment. Today a basic workstation may be no different in hardware than an ordinary desktop computer, although more powerful computers are still used for intensive calculations. Just as important as the workstation hardware is the specialized software which runs on them, and the networks which connect them to each other, to production and test areas, and the outside world.

- **Workstations** - Today an engineering workstation is merely an ordinary computer of sufficient specifications to run engineering software or to remotely access higher performance clusters. The higher end ones may have two or more processor chips, each with 6 or more CPU cores each. They can also include up to 4 graphics or parallel compute add-on cards based on graphics technology. These are used for massively parallel calculations. Typically multiple large monitors are used, and relatively large amounts of memory and hard drive storage. More moderate workstations will have specifications similar to modern gaming systems, because game graphics and engineering computations both rely on making large numbers of calculations. Even relatively powerful workstations are not expensive relative to an engineer's salary (the software they run is a different matter), so the choice of hardware will be driven more by ability to run the needed software than by cost.

- **Storage Servers** - When working on complex projects, the amount of data involved can exceed what can be stored on individual workstations, and backups should be made in case of accidental deletion or hardware failure. A storage server's main job is store the extra data where it can be accessed by anyone on the project team who needs it. That would include a history of older versions of the design, and test and simulation data, which can be voluminous.

- **High Performance Clusters** - Some types of engineering calculations require more speed than can be reasonably installed in an individual workstation. High performance clusters, or **Supercomputers** as they are also called, group many computer chips into racks with high speed data connections between them. They run specialized software designed to make use of this hardware, and the fastest such clusters represent the most powerful single computers in existence. When the need for high speed transfer between cores is not as great, the **Distributed Computing** method can be used. This harnesses a network larger single computers, or the excess computing power of a number of workstations, either off-shift or by using whatever extra processing ability is not needed by the primary user of the workstation.

- **Computer Networks** - Networks are almost universally used in modern engineering to transfer data both within a project and with the rest of the world. Since installing a network is like adding new utilities to a building, forethought should be given to making it easy to upgrade, and putting in enough network capacity that it does not need to be upgraded too often. Networking protocols and hardware change constantly like most computer-related things. Currently the most common method uses the **Internet Protocol** and routers. The protocol defines how addresses for each destination and data packets to be sent are constructed. Routers are the devices which look at the address on a packet, and send it towards the destination. There are many methods of transmitting the data between locations, ranging from Ethernet, to fiber, to wireless. In some cases it is faster and cheaper to send large amounts of data in the form of tapes or hard drives, because of their enormous storage capacity in a small package.

Computer Software

As mentioned above, engineers typically use specialized software to help with their work. The particular software will vary according the task being done. Software usually evolves rapidly, so we will discuss it in terms of categories and give some examples. If working on an actual project, a designer should find out what is the best software and most up-to-date versions available at the time. In some cases, no existing software is completely suitable, and modified or completely new software would be needed.

Analysis and Simulation Software

Historically numerical analysis relied on manual methods with devices like slide rules and tables of performance. With the advent of digital computers, special purpose programs were written in mathematically oriented languages such as FORTRAN. These performed calculations much faster than by hand, but were still limited. The processing speed and memory capacity of early computers limited the complexity of the mathematical models and how many calculations could be performed in a reasonable time. The fastest available processors in 2016, which have evolved from mainframes to supercomputers with many parallel cores, are up to a 30 billion times faster than mainframes from 50 years earlier. Desktop workstations are millions times faster than 50 year old mainframes. So the mathematical models of a design can be much more detailed and smaller time steps or more iterations of the analysis can be run. Parametric analysis allows varying parameters of the design or simulated conditions over a range of values. Since this requires multiple runs of the calculations, they have become more feasible with faster computers.

What started as individual special purpose programs is evolving into integrated general purpose suites. This reduces the need for re-entry of model data. Often the data can be used directly from the original design software, or the analysis results can be fed back to the design program directly. For some projects, custom software may still be needed where general purpose software is not adequate.

- **Numerical Analysis** - This category includes spreadsheets (for simpler analysis), general numerical calculators, such as Mathworks **MATLAB** for more complex analysis, computer algebra software, such as Wolfram Software's **Mathematica** or Maplesoft's **Maple** for symbolic problems, and more specialized programs written for particular fields. A more detailed list of **Numerical Analysis Software** can be found on Wikipedia.

- **Simulation** - This software category analyzes the behavior of a design with respect to time or changing conditions. They can cover a single type of behavior, such as mechanical stress, or multiple ones, which are called

Multiphysics tools. These can do multiple analyses in series from the same source model, or in some cases a combined effects analysis all at once. A detailed list of **Simulation Software** can also be found on Wikipedia.

Software Resources

- **Multiple Programs**

NASA Open Source Software – Repository of 240 software projects.

Public Domain Aeronautical Software - Website with many downloads of programs, source code, and documentation.

Aerospace Software Tool Library - A list of links to commercial, government, and free software, sorted by category.

Open Channel Foundation - Hosts nearly 300 mostly technical software applications, including a **COSMIC Collection** contributed by NASA.

- **Aircraft Design**

CEASIOM - Software package for airplane design. Download with registration.

- **Space Simulators**

Space Engine 0.9.7.2 - Space simulation software.

Celestia 1.6.1 - A 3D space simulator which can be used as a planetarium or for mission visualization.

- **Celestia Motherlode** - A collection of add-ons for Celestia.
- **Celestia Wikibook** - An online guide to the Celestia software.

Design and Manufacturing Software

These are the modern replacement for drafting tables. They include 2D and 3D drafting, 3D modeling, and illustration programs, and software to feed manufacturing data direct to factory machines or to vendors. Modern graphics cards and processors allow direct visualization and manipulation of the design in real or near real time. As noted above, the design and analysis software categories are becoming more integrated. Design category is also called **Computer-Aided Design** (CAD). When use of computer workstations and mainframes was new, the phrase distinguished it from the traditional drafting on paper type of design. Today design on paper is a rarity, so saying it is done with computers is mostly redundant. We group the types of software below in terms of function: drawing, modeling, and production.

2D and 3D Drafting

This category produces a set of drawings, which in turn consist of a set of lines, curves, and text or attached notes. They are distinguished from 3D models by the drawing elements existing independent of each other, and not forming more complex entities with attached non-drawing properties. Nowadays only lower-tier software such as **AutoCAD LT** or **[Solid Edge 2D** is restricted to 2D.

3D Modeling

This category defines the three dimensional shape of an object in terms of a linked set of points, lines, curves, surfaces, or volumes. In addition to the shape, a wide range of other parameters may be associated with the object. **Primitives**, basic shapes such as boxes, cylinders, or spheres, often are used as starting points, and then various operations are performed to modify or join them into more complex shapes. Wikipedia has an extensive list of **3D Modeling Software** A few examples are:

- **Autodesk Products**- Originally developer of Autocad, a 2D drawing program, this company, through acquisitions and development of new software, now has a vast range of overlapping and linked products. The tendency is to offer more integrated suites of compatible programs rather than individual ones.
- **Solidworks** suite by Dassault Systemes. - This is a high end commercial software set for design, simulation, and data management.
- **FreeCAD** is an open source 3D modeling program.

Manufacturing Software

Modern factories use extensive computer control for their operation, which in turn requires software to control the equipment. As each factory is different, the software is often customized for a given application. **Computer Numerical Control** (CNC) is the name for the machine category controlled this way. This was to distinguish it from the earlier manual control of factory equipment, and the intermediate numerical control, via stored commands, but without a computer. **Computer-aided Manufacturing** (CAM) is the process of using these kinds of machines, and the software category for producing commands and controlling the machines. Wikipedia has a very large list of **Computer-aided Technology** companies and software projects.

Software Development Software

These are tools to help make software. Many end products today require sensors, data transfer, and internal decision making and control which requires custom software to operate. Naturally enough, such software is developed on computers, using **Integrated Development Environments** (IDE) such as the **Microsoft Visual Studio** suite. When such software runs inside hardware other than computers, special test rigs and test software may be required to test the target software, and how it functions with the intended hardware. For example, a surface rover being sent to Mars is a unique item. So extensive testing would be done with software simulation and prototypes before installing it on the flight unit.

Planning and Management Software

Complex projects have to track more than just the engineering design. They have to coordinate the work of many people, do advance planning, track production and costs, etc. **Project Management Software** is designed to help with these tasks. Both project management and documentation tasks can use general office software suites, such as **Microsoft Office**, which has a compatible **Project** program. A given project can also use specialized programs for accounting, scheduling, inventory tracking, etc. Wikipedia has a very extensive list comparing various **Project Management Software** packages. There are very many other pieces of business software available. Strong consideration should be given to compatibility between programs, so that data may be moved easily between them, rather than having to convert or re-enter data.

Documentation Software

This category is used to record all the data created in a project so it can be found, shared, updated, and used.

Instrumentation and Test Hardware

Physical instruments and test equipment can be grouped into two categories: Those used in common with other industries, and those unique to space systems

Common Instrumentation and Test Equipment

Space systems projects use many of the same items as other industries to test, measure, and inspect during manufacturing, assembly, and test. Amazon's website has a large listing in that category, but there are many other sources for instrumentation and test equipment. Categories include calibration, dimensional measurements, electrical, electronics, and software testing; motion, speed, and forces; pressure and temperature, airflow and air quality, inspection and testing, light, network and cables, recording and data acquisition, weight, sound, and surface and hardness. The modern trend is to use equipment that directly feeds computer data storage, so that manual recording of data isn't necessary. Common tools, such as wrenches to remove an inspection panel, are also used, but normally those are available from production areas and don't need to be specially provided.

Special Test Equipment

Space hardware is typically exposed to two special environments. The first is launch on a rocket, followed by the conditions in space. To make sure the hardware will work properly, the hardware is subjected to a number of tests to simulate these environments. These tests require special test chambers to reproduce the conditions. Commonly used ones include:

- **Acoustic Chamber** - Rocket engines emit high pressure gas through a constriction, the engine throat, and therefore function like a whistle or organ pipe, generating huge amounts of sound and vibration. The sound portion is tested in an acoustic chamber with powerful speakers, which play a noise spectrum matched to the launch vehicle the hardware will ride on.
- **Shaker Table** - The high speed flow from the rocket engines, and air flowing past the rocket in flight is turbulent, generating physical vibration in the vehicle. This is distinguished from sound that travels through air. Vibration is simulated by a table that holds the hardware the same way it is held for launch. The table is moved in all directions with powerful pistons, springs, and unbalanced masses, to reproduce the levels the space hardware will experience.
- **Zero-G Deployment** - Spacecraft often have solar panels, antennas, and other items that are folded to fit in the payload space of a rocket, then unfolded once in space. The unfolding happens in zero gravity, and this is simulated by doing it sideways, with counterweights to remove the weight from the joints and mechanisms.

- **RF Chamber** - Most spacecraft communicate through radio frequencies (RF) and antennae. Antenna operation and links to the rest of the spacecraft are tested in am RF-shielded chamber and separate transmitters that simulate ground stations.

- **Thermal-Vacuum Chamber** - The space environment is usually in vacuum. Hardware is subjected to cold from the cosmic background near absolute zero, and heat from the Sun, which is more intense above the atmosphere, or if the mission goes closer than Earth to the Sun. Since vacuum does not allow heat transfer by conduction through the air, different sides of a spacecraft can be hot and cold at the same time. These conditions are tested in a large vacuum chamber, which is provided with cooled walls and intense lamps to simulate the cold and heat conditions.

Beyond these devices, which are commonly used for whole spacecraft, special purpose equipment may be needed for particular instruments. For example, the **Chandra X-Ray Telescope** needed a 300 meter vacuum tunnel to test the X-Ray optics from an optically distant source.

Section 1.7 - Engineering Specialties

In **Section 1.5 - Systems Engineering** we discussed methods to coordinate the work across a large project. Complex space projects require more knowledge and experience than any single individual can have. Such projects need teams of people not only because there is a lot of work to do, but also because each person supplies a different set of skills. For the design portion of a project, the field of engineering is divided into a number of **Branches** of knowledge, with specialists who concentrate on wider or narrower portions of it. The specialists address different areas of design, and different areas of application to a given project.

Much of the actual work for current space projects happens on Earth, in places like offices, factories, launch sites, and control centers. Building and operating those locations uses much of the same engineering knowledge as any other large project on Earth. For the in-space segment of projects, many of the specialties are still relevant. For example, nearly all space hardware has structural parts, and their design is the province of mechanical engineers. In the future, more production and construction will happen in space. This is in contrast to today, where mostly finished hardware is launched from Earth. So the importance of additional fields like mining and industrial engineering will increase, with suitable modifications for operation in space. In the farther future, extremely large projects described as **Astrophysical** or **Planetary Engineering** are possible. Examples are **Terraforming**, making a body more Earth-like, or changing the orbit of a large asteroid. Such very large projects are mostly speculative, except for the human-caused 43% increase in CO_2 in the Earth's atmosphere. This is anti-terraforming our planet - making it less Earth-like than its original state. Large-scale space projects are one way to correct this problem if it becomes too severe. Very large projects like these are not yet organized as a distinct specialty, but would include knowledge from many areas of science and engineering.

One branch in particular, Aerospace Engineering, is concerned with the design and construction of vehicles and hardware that travel to and operate in air or space. We discuss it first, because it is nearly always involved to some degree in space projects. It should be remembered that aerospace engineers are part of a team, and not the only skill required. Although engineering has many branches, they all rest on a common foundation of the sciences and mathematics, so there is some overlap between them. For example, aerospace and mechanical engineers analyze loads on space vehicle structures the same way civil engineers analyze them for ground structures. Where they differ is in what materials are used, where the loads come from, and their operating environment.

The remainder of this section lists other major engineering branches with some relevance to space projects. We will not go into great detail about them, but a space systems designer, whether generalist or specialist, should at least know what other areas exist besides their own. They can then find detailed information on a topic, or find specialists as needed, when it is beyond their own area of knowledge. For those who want to learn more about a particular area, one place to start is the Massachusetts Institute of Technology (MIT) **OpenCourseWare** website, which has an increasingly large collection of college level open source course material available (about 2250 so far). We list a number of their courses below, but they are not an exclusive source. Additional information can be found through the links below, the **References Section** at the end of this book, and the huge number of books written on engineering topics.

Aerospace Engineering

This is the primary field concerned with the design of systems which operate in the atmosphere and space. It is further divided into **Aeronautics**, having to do with flight and operation in an atmosphere, and **Astronautics**, having to do with travel and operation in the space environment. The latter is the primary subject of this book. An introductory course on this subject is **Introduction to Aerospace Engineering and Design (MIT)**. Aerospace engineering work is also divided into specialties according to how travel is accomplished, the environment that systems travel through and operate in, and the internal subsystems which make up vehicles and other space hardware. These specialties include:

Dynamics

In general physics and engineering, dynamics is the evolution of physical processes with time. General courses on this subject include:

- **Dynamics (MIT)**
- **Aerospace Dynamics (MIT)**

Within aerospace engineering these particular areas of dynamics are important:

Aerodynamics - uses knowledge of fluid and gas dynamics as applied to the interactions of atmospheres with primarily solid objects such as vehicles. Vehicles going to space, or returning from it, must traverse the Earth's atmosphere. Some transport methods also actively use the atmosphere, and some destinations in space have their own atmospheres to be navigated. A combination of high velocity and atmospheric density can create large forces and high temperatures. These must be accounted for in design. Natural movements of atmospheres, such as wind and **Gravity Waves**, fall under the operating environment for aerospace systems. Relevant online courses include:

- **Aerodynamics (MIT)**
- **Aerodynamics of Viscous Fluids (MIT)**
- **Compressible Flow (MIT)**

Astrodynamics - which is also known as **Orbital Mechanics**, is the application of ballistics and **Celestial Mechanics** to practical problems. Celestial mechanics is the branch of astronomy that deals with the motions of natural objects in space. Space systems that are not using propulsion, nor interacting with atmospheres or magnetic fields, follow the same motions as natural objects. The path a system follows, called a **Trajectory** or **Flight Path**, can be set up with periods of natural motion under gravity, and periods of active propulsion. The natural motion can include coasting between objects, and **Gravity Assists**, which are close passes of a larger object to affect velocity and direction. Relevant online courses are:

- **Astrodynamics (MIT)**

Structures and Mechanisms

Structures and mechanisms are the load bearing and mechanical parts of an aerospace system. The primary structure carries the main loads from gravity, acceleration, aerodynamic forces, etc. Secondary structure holds equipment items in position, which are lesser loads. Mechanisms are the moving parts of the system, such as joints and actuators. Typical mechanisms do steering for a rocket engine, or unfolding and pointing of solar arrays. Aerospace engineers can specialize in this design area, but they have significant overlap with **Structural** and **Mechanical Engineers**, who also deal with load-bearing structures. The specific knowledge for aerospace systems involves operating conditions like high accelerations, vibrations, large temperature ranges, and exposure to vacuum. It also involves specialty materials to save weight. These are used for performance and program cost reasons. Underlying knowledge for this system includes **Materials Science**, which covers the relationship between the structure of materials at atomic scale and their larger scale properties, and the selection of materials for particular applications. It also includes **Solid Mechanics**, which is the behavior of continuous solid matter under external actions, such as forces, temperature changes, or applied movements. Modern design of structures and mechanisms is usually performed with computer-aided software, and increasingly is integrated in process. This goes from defining the shape of parts, to analysis and simulation, optimization, and then submitting design files to computer-controlled factory equipment to produce them. Once designed, a structural **Test Article** is often built to prove the physical version can withstand all the design conditions. Relevant online courses for this subject include:

- **Structural Analysis and Design (MIT)**
- **Computational Methods in Aerospace Engineering (MIT)**
- **Plates and Shells (MIT)**
- **Computational Mechanics of Materials (MIT)**

Power and Electrical Systems

This system is concerned with the supply of power, and electrical systems such as heaters and motors, with respect to aerospace systems in particular. It overlaps with **Electrical Engineering**, which is the more general subject (see below). Some space equipment needs high levels of power for a short time, like during launch of a chemical rocket. This can be produced by an **Auxiliary Power Unit** which uses a turbine and chemical fuel. For longer-term needs, solar arrays and sometimes nuclear sources are currently used. Selection of power sources for space projects requires understanding their operating environment, need for long-term operation without maintenance, and other special conditions. Once generated, the power may be stored temporarily in devices like **Batteries**, and in all cases is distributed through cabling, with fault protection, switching, regulation, and control.

Propulsion Systems

Propulsion Systems in general include a source of power, and means of converting this power to propulsive force. The purpose of these systems is to move people or goods over some distance, usually as part of a **Vehicle**, an artificial carrier. **Space Propulsion** is the methods of propulsion useful for space projects. Some of them only work in space. Others work in an atmosphere or on the surface of an object, and are then similar to methods used on Earth. A wide variety of space transport methods are listed in **Part 2** of this book. Not all of them involve vehicles with propulsion systems. For the ones that do, chemical, electrical, nuclear, and other power sources are used. A variety of engineering specialties are therefore involved in their design. Aerospace propulsion engineers have specific experience in one or more of the space transport methods, as opposed to transport methods used on Earth. Relevant online courses include:

- **Introduction to Propulsion Systems (MIT)**
- **Rocket Propulsion (MIT)**
- **Space Propulsion (MIT)**

Thermal Systems

Thermal Systems have the function of keeping all parts of an space system within acceptable temperature ranges. This includes passive methods that work because of their inherent properties, such as insulation, coatings, thermal couplers and isolators, reflectors, and radioisotope heaters. Active methods use devices like electric heaters, heat transfer fluids and radiator panels, mechanical louvers, and thermoelectric coolers.

The air and space environment, and the operation of internal systems, can generate wide variations of temperature. The main natural source of heat is the Sun, which is nominally 36% more intense in space due to lack of atmosphere, but this varies according to distance. Secondary heat can come from reflection or thermal emission from nearby large objects. The main source of heat loss is the **Cosmic Background**, which is in all directions in the sky behind individual foreground objects. It consists mainly of the **Cosmic Microwave** and **Cosmic Infrared Backgrounds**. Their combined effective temperature is about 3 K above absolute zero, or -273 C. This is much colder than most space hardware, so more heat is lost to the background than gained from it. Besides the natural environment, operating hardware that makes up a space system can generate heat or be very cold. For example propulsion systems can do both - generating large amounts of heat in a rocket engine, while having cryogenic temperatures in the propellant tanks.

Thermal control engineers analyze spacecraft environments and operations to determine what temperatures will occur. If hardware plus natural heat results in an unacceptable temperatures, then thermal control is required, and these engineers help design the solutions. **Heat transfer** in general is a topic of physics, and is addressed as a design task by mechanical engineers (see below), but specific problems and conditions in aerospace require specialized solutions. Some relevant online courses include:

- **Thermal Energy (MIT)**
- **Radiative Transfer (MIT)**

Control Systems

Control Systems Engineering is the specialty that applies **Control Theory** to design active systems with desired behaviors. Since most space systems have active components, they also need control elements. These include sensors and instruments to detect the current status, devices to transmit, store, and process the data thus generated, methods to generate the appropriate response, and to transmit commands to other parts of the system. The commands are implemented by actuators, such as a valve in a rocket engine that controls fuel flow. Typically control systems operate in a **Closed Loop** fashion. This is where a cycle of detecting status, processing data, issuing commands, and then detecting the new status repeats multiple times. The manufactured parts of aerospace control systems are are often called **Avionics**, short for "aviation electronics", even if used in space systems. **Human-in-the-Loop** control systems include humans as part of their operation. For example, airplane pilots use their eyes, brains, and hands as part of the control loop.

Control systems are used in everything from traffic lights to chemical plants, so the subject is taught to all kinds of engineers. Aerospace controls specialists deal with the particular problems of air and space operations, such as control of a rocket in flight, or a space system's robot arm. Some relevant online aerospace courses include:

- **Principles of Automatic Control (MIT)**
- **Feedback Control Systems (MIT)**
- **Prototyping Avionics (MIT)**

Space Environment

The **Space Environment** has a number of significantly different conditions than found on Earth. These must be accounted for in space systems design, and so specialists in environment effects on systems and people are needed. The space environment includes all external factors that can affect a system - such as free fall (zero gravity) or low gravity; vacuum, rarefied, or different atmospheres, which can cause drag, erosion, or electrostatic charging, and wide temperature fluctuations. Particular hazards to humans and space hardware are unique to space. They include:

- **Meteoroids** and **Space Debris** - Natural and artificial objects that, due to their high relative velocity, can cause damage on impact.

- **Space Radiation** - Particles of high enough energy to damage people and equipment. On the Earth's surface we are sheltered by the magnetic field and atmosphere from the naturally high radiation levels that exist most other places in the Solar System. Sources include **Cosmic Rays**, high energy particles and photons from the Sun, and trapped **Radiation Belts** around objects with magnetic fields. Artificial nuclear devices used in space can increase radiation levels, and thicknesses of mass or magnetic fields can shield from it.

Study of the space environment is part of the science of Astronomy, especially **Astrophysics** and **Planetary Science**. These fields have benefited greatly from space science missions, which have allowed measuring the environment directly. Environment effects engineers often come from the sciences first, then apply their knowledge to space projects. Alternately, they start from an engineering specialty, then add the relevant scientific knowledge, often at a graduate level.

Life Support Systems

The space environment is generally hostile to humans, and life in general. For space projects that involve living things, a **Life Support System** is then needed to provide suitable conditions. Basic structure contains an atmosphere, and thermal control keeps temperatures in a habitable range. These were covered previously. Beyond this, humans need the right atmosphere mix, water, and food. Because of the closed environment, liquid, solid, and gaseous wastes (including especially CO_2) must be removed, both from people and other system operations, and microbes controlled. In free fall conditions air does not circulate by convection, so circulation fans are needed to prevent "dead zones" where harmful concentrations of gases can accumulate. Biological systems can be used to recycle wastes and supply food, like they do on Earth, but this is still in early research. Most life support systems to date have single-use supplies of food, and limited recycling of other materials.

Engineering of life support systems is cross-disciplinary, involving both biology and mechanical systems. Specialists therefore come from areas like Bioengineering and Mechanical Engineering (see below). They typically learn about the specific design of life support on the job, because there are so few examples of life support for space that it has not developed an educational path yet. Related work is done for airplanes, high altitude climbing, and working underwater. Life support systems can be large enough for a number of people, or small enough for a single person, as in a space suit.

Human Factors

Human Factors, also is the aspect of design that takes account of the interaction of a system and the people who use it. People can't be designed the way a piece of hardware can, so the system design has to accommodate their capabilities and limitations. The people may be passive, as in airplane passengers, or active, as in the flight crew. Subject areas in this field include physical, cognitive, and organizational interactions. Because of the unique conditions in space, like having to do repairs while in zero gravity and a bulky space suit, or working with a large support team on the ground, the various subject areas have assumed importance in aerospace. Human factors also includes topics like how to design control inputs and information displays for zero gravity or high acceleration, and how to maintain crew training on a long mission.

The roots of human factors design extend as far back as people have made tools, since the tools must fit our hands, and the strength we have to wield them. The modern development draws from disciplines like psychology, engineering, biomechanics, industrial design, physiology, and anthropometry. A relevant online course is:

- **Human Factors Engineering (MIT)**
- **Human Control of Automated Systems (MIT)**

Simulation and Test

Aerospace systems are often complex, and therefore expensive. The environment and operating conditions in which they operate are different and more severe than those normally found on the Earth's surface, and failures can be catastrophic. So to ensure basic functionality, safety, reliability, and meeting design requirements, the **Simulation** and **System Test** specialties have developed.

Simulation includes early mathematical and computational modeling, and physical scale models or functional simulators. They reproduce important aspects of a system, but do not use the actual hardware and software products. Testing uses actual materials, components, subsystems, up to completed products. Test can be in simulated environments, like a **Vacuum Chamber**, or in the actual operating environment. For complete aerospace vehicles, the latter is called **Flight Test**. For very large and complex systems, such as the International Space Station, there was no way to test it as a whole. Instead its parts were tested on the ground, extensive analysis was performed, and the ability to repair and update parts as needed had to suffice.

The **Hardware** and **Software** test planning should ideally start early in a project. Specialists in these areas are often drawn from the respective design fields, and from Systems Engineering experience.

Other Engineering Specialties

The following are major conventional divisions of the engineering field into specialties, listed alphabetically. Knowledge in general does not have such divisions, they are made by humans for historical and teaching purposes. A given engineer may have knowledge and experience that spans across multiple specialties or is concentrated in a narrow area within only one. Where we listed individual courses for specialties within aerospace engineering above, the larger engineering fields tend to have whole academic departments or even entire institutions devoted to them.

Bioengineering

Biological Engineering, or "Bioengineering" applies knowledge from the biological sciences towards satisfying human needs. This includes producing food, materials, energy, maintaining human health, and the natural environment. As an engineering field it has developed rapidly since about 1960, because of the increased understanding of genetics, and development of tools to manipulate it. The **MIT Department of Biological Engineering** lists a number of open courses on this subject.

Agriculture

Agricultural Engineering is the subset of Bioengineering concerned primarily with food, wood, fibers, biofuel, medicinal, and other material products produced on land. As a human activity, **Agriculture** extends back to the origin of civilization. As one based on scientific knowledge it has greatly improved in the last 200 years, and as an engineering field extends back to about 1900. Growing plants in space is at an early experimental stage. This is expected to increase greatly in coming decades, since growing food and other products, and production of oxygen as a by-product, would greatly reduce supplies needed from Earth.

Biomedical Engineering

Biomedical Engineering applies engineering principles and design concepts to medicine and biology for healthcare purposes. It is a relatively new field with a heavy emphasis on research and development. For space projects, it is applied to keep crew and research animals in good condition, and also a subject of research on the effects of space conditions.

Chemical Engineering

Chemical Engineering is the application of physical and life sciences, and applied mathematics and economics, to produce, transform, transport, and properly use chemicals, materials, energy, and useful products. It is a broad field with many specialized branches, and typically has a full set of courses at schools such as **MIT's Department of Chemical Engineering**. Historically, the relevance to space projects has been supplying propellants for rockets, and the alloys and other materials to build them and their payloads. In the future, extracting fuel and other products from space locations will require modified versions of chemical plants.

Civil Engineering

Civil Engineering deals with the design, construction, and maintenance of the built environment, including roads, bridges, canals, dams, and buildings. As an empirical practice it is as old as **Civilization** (both words derive from the Latin *civitas*, or city). As a scientific and technical field it began around 1800. For space projects to date it is mostly related to factory locations and launch sites on Earth. In the future it will be applied to construction at locations beyond Earth. As one of the older engineering fields, it typically has a full set of courses, such as at **MIT's Department of Civil and Environmental Engineering**. Branches of civil engineering include Environmental, Geotechnical, Structural, Transport, and Water Resources Engineering. Since all the materials used to build with must be obtained from the natural environment, it is closely related to Mining Engineering (see below).

Electrical Engineering

Electrical Engineering deals with the applications of electricity, electromagnetism, and electronics. It became an identifiable field after about 1850. Most space systems have electrical components, and the specialized aspects are noted above under aerospace engineering. More general electrical engineering is used today in supplying power to offices, factories, and launch sites for space projects. It is also used in computer hardware, software, and communications networks for development and operation of space systems. Future uses include electric launch and space propulsion methods, and solar power delivery from orbit. As one of the larger fields, it has a range of courses, such as at **MIT's Department of Electrical Engineering and Computer Science**

Industrial Engineering

Industrial Engineering deals with the optimization of complex processes or systems, and improvement of productivity, quality, efficiency, and profitability of businesses. Traditionally it encompassed planning of industrial production systems (factories), but has broadened in scope to a wide range of complex operations. Industrial is distinguished by being concerned with the whole business and supply chain, where manufacturing engineering (part of mechanical below) is concerned with individual production machines and tasks.

Mechanical Engineering

Mechanical Engineering applies the principles of physics and materials science for the design, analysis, manufacturing, and maintenance of mechanical systems. Specialized mechanical systems for space are listed above under aerospace engineering. Other specialties include acoustical, manufacturing, thermal, sports, vehicle, power, and energy engineering. Mechanical systems such as turbopumps and valves are core elements of rocket engines. Wheels, suspensions, drive motors, and robot arms enable rovers to perform complex tasks on planet surfaces. As a modern technical field, mechanical engineering has grown since about 1800, with the development of **Machine Tools** and engines to drive them. As one of the largest engineering fields today, it has a very large range of courses, such as at **MIT's Department of Mechanical Engineering**

Mining Engineering

Mining Engineering is concerned with extracting and processing minerals from the natural environment. Modern civilization uses vast quantities of materials. So this field of engineering, and closely related sciences, have developed to support those uses. At present, mining on Earth is not directly related to space projects, as raw materials are normally processed by intermediate factories and plants before being used to build hardware. This may change with the development of integrated and automated production systems. In space itself, extracting materials is of great interest to avoid the high energy and financial cost of launch from Earth, and making available new sources of materials and energy. Mining in space is still in the research and early development stage.

Nuclear Engineering

Nuclear Engineering is concerned with applications of nuclear processes like fission and fusion. These processes can release very large amounts of energy. On Earth the primary use is for power generation in **Nuclear Reactors**. Some space missions today carry sources based on radioactive decay. In the future fission and possibly fusion reactors can supply power when solar is insufficient. Nuclear engineering deals with high energy radiation, which also is part of the natural space environment. Although a fairly new field, dating to the 1940's, it has developed extensive specialized knowledge, such as courses from **MIT's Department of Nuclear Science and Engineering**.

Software Engineering

Software Engineering is the systematic application of engineering methods to the design, development, and maintenance of **Software** - the changeable instructions and data that computers use. This field is young, only developing from about 1950, but has rapidly grown due to the rapid advance of computer hardware and electronics in general. Software is an integral part of modern space projects, from initial concept formulation using **Productivity Software**, to collection and analysis of mission data and control of remote spacecraft. Essentially all engineers use software, and many develop software as part of their work in other specialties. However creating complex and reliable software, where the consequences of errors mean loss of life or expensive systems, requires teams of specialists dedicated to the task. Software engineering is closely related to **Computer Science**, and their teaching is often combined in one set of courses, like at **MIT's Department of Electrical Engineering and Computer Science**.

Section 1.8 - Organization and Economics

The Systems Engineering approach described in **Section 1.5**, along with the tools and engineering specialties in sections 1.6 and 1.7, cover parts of how design gets done. Although not strictly part of the design process, a large and complex space project also needs an effective **Organizational Structure** to function. An organization allocates tasks, coordinates the work, makes decisions, supplies the various resources and inputs needed, and provides support to the core design, production, delivery and operations tasks. Individual projects may be part of a larger program or ongoing business, and organizations must then sequence and allocate between them.

Economics is the social science that studies production, distribution, and consumption of goods and services. Space projects are then an economic activity, and part of a general economy. Large projects need a source of necessary resources, which is often in the form of money. The tools and methods of economics are therefore used to analyze, justify, and obtain the needed money and resources. Organizations and economics are extensively studied at business schools, such as the over 100 courses at **MIT's Sloan School of Management**, and there are many textbooks on these subjects.

Organizational Structures

There are a number of organizational structures that have been used or are possible. Which ones are best suited to a project or program depends on the complexity, duration, and life-cycle stage of the tasks. Structures may be changed or replaced over time as the needs change. Humans are social entities, and their individual needs, desires, and motivations differ. This has to be considered along with technical questions like what kind of desks and workstations to use. Membership in an organization is usually not permanent, either because the project's or the individual's needs change. Continuity of a project has to be supported when this happens.

Structural Types

For a given project, you can adopt an existing structural type. For an ongoing program or business that structure may already be in place. You can then modify the existing type to accommodate the specific needs of the project. Alternately you can design a custom organization in the same way as you design a custom piece of hardware. The choice will depend on the potential gains from customization, versus the lower adoption cost of an existing type. Existing structural types include:

- **Entrepreneurial** - One or more founders or leaders make decisions, and communications are direct person-to-person. Staff are added on an ad-hoc basis, and tasks are not rigidly defined. This type is suited to new projects at an earlier stage of their life cycle, where many decisions have not yet been made, and change is rapid.

- **Bureaucratic** - This may be thought of as a machine made of people. The structure has a degree of standardization, with defined roles and responsibilities, a hierarchical (pyramid) structure, and respect for merit. It is suited to larger and more complex organizations. Since one person can't comprehend all the details or have time to make decisions for a large organization, information flow and instructions are compartmentalized through the pyramid structure, with a given person responsible for actions at their level, and one level up and down. Rules and standards make tasks routine, so decisions are reduced to exceptions or changes. The accumulated structure and rules makes this type harder to change when needed.

- **Networked/Consensus** - In the previous two types, decision-making is concentrated in a few leaders. In a network or consensus structure, decisions are reached by multiple interested parties. Communications flows to and from any point in the network as needed. This is suited to more rapid and ad-hoc change, but increases effort to make decisions because more people are involved. However, increased information inputs can lead to better decisions, and wider participation can lead to better acceptance and enthusiasm.

Structures are not usually pure examples of any of these types. For example, bureaucratic ones may have boards or committees which reach decisions by consensus, and organization members may communicate through an informal network in addition to a formal pyramid structure. The parts of a structure may be defined by specialized functions, such as production, marketing, staffing, and accounting. It may also be divided by location or product, where each part includes multiple functions. Again, mixed examples are possible for a given organization, with some parts functional and others location or product-based.

Designed Organizations

An organization capable of designing and building complex space projects is usually itself a complex entity. Therefore systems engineering methods can be applied to the design of the organization itself, as well as to the space systems they produce. In this approach, the organization is treated as a system, where inputs like staff time, office buildings, and factory buildings produce outputs like launch vehicles and spacecraft. By comparing the outputs to the inputs, you can decide if a given project is worth doing. The internal functions and flows within the organization determine the relationship of inputs to outputs, and so are subject to analysis and optimization.

Designing an organization this way is not often done for several reasons. First, organizations often have a history, and existing structures and methods in place for how they operate. These may have developed before they started on space projects, and even before the systems engineering method existed. If existing organizations and methods have worked well enough, the need to optimize or update them may not be felt. Second, people are often resistant to change, especially when that change involves re-designing or eliminating their own job. Some people are attracted to the power or money that comes from a position within an organization, and so oppose any change that would displace them from that position. Third, many people in a high position believe they know better how to organize and run things. This results partly from the well known psychological cognitive bias of **Illusory Superiority**, where people in general misjudge their ability relative to others. It also partly is due to the belief that reaching a high position by itself demonstrates superior knowledge and ability. Lastly, there is the fear that any changes will break how the existing organization operates. This is a fallacy because organizations in general are capable of doing things no single individual can do, like design an entire airplane or computer operating system. Therefore as a whole they should be more capable of designing the organization itself than any single individual within it. Nonetheless, these various barriers exist and should be recognized.

If an organization is new, or being organized for a specific project, most of these barriers cam be overcome. Then a rational application of the systems engineering process can find the best organization structure. This can be one of the existing types, since they exist for good reasons. But choosing a structure by design is better than choosing one by personal preference or random accretion, since it is more likely to produce a better result. Another way to implement change is to use an outside organization to design the new structure, present their results, and then have the old organization adopt it as a whole. Approached in this way, individual objections can be overcome, but it can also induce loss of staff beyond that caused by increased efficiency.

Project Management

Project Management is the discipline of initiating, planning, executing, controlling, and closing the work of a team to reach specific goals and success criteria. A **Project** has a defined goal and end point, whereas a **Program** is a larger effort that may include multiple projects. Programs may be open-ended with no definite end point, such as "explore the Solar System". Programs use more permanent organizations such an agency, research institute, or business enterprise. Individual projects can draw from these organizations, and later return staff and other resources to apply to the next project.

Historically, skilled and experienced individuals such as architects, engineers, and senior craftsmen led large projects, which were usually in construction. By the 1950's, project management had become a distinct discipline from these skilled fields, with its own tools and techniques. There are a number of approaches to managing a project, including applying systems engineering to the whole project. The rapid growth in complexity and scale of projects has driven the creation of new methods to manage them, and this is an ongoing process today. New tools like computer networks and software enable some of these new methods. See Wikipedia's **Outline of Project Management** for a more detailed overview of the field

A **Project Manager** is a person given responsibility for part or all of a project's execution. Traditionally managers are placed on top of an organization structure, but that is a matter of history, not necessity. They combined several functions such as planning, decision making, and coordination, in an era when communication was more difficult. Modern computer networks allow everyone on a project to have immediate access to all project information, so the need for top-down hierarchies is reduced. Quite large and complex projects, like the **Linux** operating system, have been developed mostly with self-directed work. However, many space projects require dedicated specialists, large buildings, and other resources that are difficult to obtain on individual initiative.

Typical project management processes are:

- **Initiating** - which includes identifying needs and requirements, reviewing existing operations, financial analysis, stakeholder analysis, and establishing a project charter
- **Planning** - of time, cost and resources for the project tasks. This includes a breakdown of deliverables and organizing a logical sequence of the work.
- **Production and Execution** - consists of following the plan, allocating staff, outside supplies, physical space, and other resources as required.
- **Monitoring and Controlling** - includes observing production and execution to identify and correct problems, and measure performance and variances from the plan.
- **Closing** - includes formal acceptance of the project outputs, ending contracts, reassigning resources, and archiving data and knowledge gained.

Project Funding

Large projects need a way to obtain all the resource inputs needed for their completion. An existing organization or program may have some of these resources, like land and buildings, in place already, but nearly always some additional ones need to be supplied from outside the project. In modern economies, **Money** is a generally accepted medium of exchange, which can be traded for all these resource types. This simplifies the resource provision task to obtaining enough funding in monetary form. There are a number of different sources and ways to obtain project funding.

For space projects, the sources can be divided into three main sectors - government, business, and social - which relate to major parts of economies in general. We will list them individually, but a mixture can be used for a given project. As examples, governments often hire businesses to build parts of space projects, businesses use government-owned launch sites, and partnerships between sectors are common. The differences involve sources of funds and objectives, but not as much the technical issues. A rocket engine works the same way no matter who pays for it. What kind of rocket engine gets built, however, may depend on available funds and who is the source.

Government Funding

Governments were the first large source of funds for space projects, growing from military uses like ballistic missiles and spy satellites. This was followed by scientific and earth observation missions such as weather satellites. These uses are intended for the general public benefit, and so obtain their funds through taxation or appropriation. Governments decide how to apply the funds to projects by political processes. This is typically by means of annual budgets, where space projects are part of more general government funding. Although theoretically for the public good, in practice other factors enter into political decision-making. Typically space projects are carried out by a mixture of internal government staff and facilities, and outside purchases of goods and services.

Business Funding

Businesses normally operate to generate profits for their owners. Early space projects were high risk, and so were mostly carried out by governments. Businesses were hired to carry out parts of these projects because of special skills or production capacity the government did not have. Since government guaranteed payment, this lowered the risk and businesses were willing to participate. Communications emerged as a profitable space business because information has no mass, and so the high cost of space transport did not affect that business as much. Space industry beyond government projects is now the dominant economic sector. In the future, additional types of space businesses are expected to grow as transport costs come down. Businesses obtain their funds from investors, retained previous profits, and finance on the open market.

Social Funding

Social funding comprises non-profit foundations, research organizations, private donations, and volunteer efforts. Large terrestrial telescopes are an example of projects funded by a mixture of social and other sources. This type of funding is at a relatively small scale compared to government and business.

Financial Analysis

Whatever the source of funding, large projects have to justify themselves to be approved, estimate the funding required, optimize the use of funds, and implement the acquisition and distribution of funds during the project. Various tools from economics and **Accounting** are used for these purposes. Money is often used as a common measure to compare disparate resource and output types. However it should not be the sole measure to evaluate projects. Measures of Effectiveness, as described in Section 1.5, are a way to evaluate non-monetary factors like performance and safety along with monetary ones like cost and revenue. These kinds of measures can also be applied at the organization level.

Finance is the science of money management. Among its other uses, money is a measure of value to the end users of a project, and a measure of the resources a project will consume. The concepts and formulas from **Economic analysis** are useful in project decision-making and optimization. If there is too little value in relation to resources, a project may never happen at all. Resources include human labor, physical resources, and intangibles such as rights to use something. The resources to run the organization itself is often a major part of total project resources. So organizational efficiency is important to reaching desired program cost and schedules.

Project Finance is obtaining money for a project from other than internal sources of the sponsor. This is often necessary for new organizations, and for large projects which exceed the scale of previous organization efforts. Financial estimates are developed before money is actually required to determine if a project is worth starting, and how much funding will be required. Actual spending during a project is tracked to compare progress to estimates, in order see if additional funds are needed or if there is surplus to apply elsewhere.

Life Cycle Cost Analysis

Life Cycle Cost Analysis is an economic method that accounts for all costs of a project, from initial concept to final disposal. It parallels the general Systems Engineering method of considering the entire life cycle of a project from start to finish, but focuses on the financial aspects. The method applies some basic financial concepts:

- **Time Value of Money** - Human life is finite. Therefore the amount of time you can expect to use resources, or their equivalent in money, is shorter at a future date than it is in the present. Changes between now and a future date may affect the desirability of any given item. For example, an old computer that may have been useful at one time has become less so due to changes in technology. Inflation can also affect the future value of money itself. For these reasons, humans attach greater value to something now than in the future, a phenomenon known as **Intertemporal Choice**

- The **Interest Rate** is a measure of how much more value the present has over the future. It is the percentage increase per time, usually years, needed to equate a future value to a present one. In other words, how much larger a future value do you demand to delay a present use? Reversing the time order, a future value is reduced by a percentage per time, called the **Discount Rate**, to equate it to the present.

Costs and benefits of a space project may extend for a number of years, and costs generally happen before benefits. Therefore both are equated to **Present Value** by applying the discount rate. Which discount rate to use depends on who is providing funding. A government generally has a lower one than a business because they assume a longer time horizon to collect benefits.

Financial Estimating

Future costs and benefits are not known exactly, especially when a project includes developing new technology. Since these predictions are often key in deciding if a project happens at all, or which design options are used, it is very important that they have as sound a basis as possible. Therefore various estimating methods have been developed to predict the financial flows for a project. They vary in detail and accuracy based on stage of the project. Regardless of which method is used, some margin or range should be included in any estimate to allow for errors in estimating, or unanticipated costs. Methods include:

- **Historical Costs** - This is based on similar past projects, and uses units like number of person-hours, drawings, or lines of software. These are multiplied by historical rates for the given parameter

- **Parametric Estimates** - These are formulas for an entire finished item using multiple parameters such as mass and stress level, and factors such as "simple design" or "complex design" to modify the basic values.
- **Bottom-Up Estimates** - This approach makes estimates for individual parts of a project at a detailed level, then summing them to get the total cost.
- **Supplier Quotes** - A supplier or sub-contractor who makes a definite quote on part of a project replaces an estimate with a specific number

Large space projects typically have many cost components, and use specialists for cost estimating and tracking, with tools like spreadsheets to integrate the data into useful summaries.

NASA has a **Cost Estimating Handbook** developed internally for space projects. The International Society of Parametric Analysts has a **Handbook** for more general projects.

Section 1.9 - Existing Programs

A space systems designer should be aware of existing projects and programs in order to avoid duplication of effort, know the current state of the art, serve as a baseline technical level to improve upon, and as a comparison point to justify the need for new projects. Projects and programs can be categorized by **type of funding** - government, commercial, or not-for-profit sectors; **time** - past, current, and proposed future; and by **type of project** - science and exploration, communications, tourism, mining, etc. There is significant overlap and interaction between the various sectors and projects. For example, government programs procure equipment and services from commercial entities, and contribute funding to not-for-profit programs. Another overlap is that launch vehicle programs provide transportation for different payload projects.

Wikipedia has extensive article lists for **Spaceflight** in general (see also the Category, Portal, and WikiProject links on that page), more specific **Lists and Timelines**, and categorized **Spaceflight Navigation Boxes** (used in articles to link to related articles). These can be used to get general background on the various projects and programs. According to the **UCS Satellite Database**, as of August, 2015, there were an estimated 1305 operational spacecraft in Earth orbit. There are also approximately 20 active **Solar System Probes** beyond Earth orbit, and 6 spacecraft at **Sun-Earth Lagrangian Points**. The latter does not include the many natural object at various Lagrangian Points. In addition to active spacecraft, there is a large amount of human-made **Space Debris**, nearly all in Earth orbit. They can be considered either a hazard or a resource. Finally there are all the natural objects of the Solar System, which can be considered as destinations for exploration, resources to be used, or hazards to be dealt with, such as radiation belts and sources.

As of 2015, there are seventy civilian **Government Space Agencies** which pursue their own space projects and programs, or participate in some way. There are also a number of military programs and projects, but due to secrecy, their number and activities are less certain. The following sections cover some of the more important programs, but are not exhaustive. It mainly lists those that are directly related to space. Other fields, for example **Materials Science**, can develop new structural alloys that can be used in space, but that is not the field's main goal. For the many such fields, which are less directly related to space, we will not list them here. If you need information about these subjects, there are many online, library, and academic sources which can be used as starting points to research them. The other fields are important, however, in that general improvements in science and technology affect what you can do for space projects.

Government Programs - International

We will start with the major national and international agencies outside the United States, approximately in order of annual budgets. It is approximate because currency exchange rates fluctuate, and so do agency budgets in terms of their own currency.

ESA

The **European Space Agency** (ESA) is an intergovenmental organization of 22 member nations. It's budget in 2015 was 4.43 billion EUR, or $5.15 billion dollars. It was established in 1975, and operates a spaceport in French Guiana, designs launch vehicles, a number of independent spacecraft, and participates in international projects and programs. The spaceport is located outside Europe for safety and launch performance reasons. Member nations typically split their space program funding between ESA and domestic projects.

Roscosmos

Roscosmos State Corporation is the current space science and aerospace agency for Russia. It's budget in 2015 was 186.5 billion RUB, or $2.4 to 3.7 billion in dollar terms. The dollar number is variable because of severe exchange rate fluctuations. Prior to the dissolution of the Soviet Union, the **Soviet Space Program** had a long history starting in the 1930's, with notable theoretical work starting before that. The program had many notable firsts and continues today in areas like participating in the **International Space Station** (ISS). During the Soviet era, military and civilian space activities were not separated, and were very secretive.

JAXA

The **Japan Aerospace Exploration Agency** is the national aerospace agency for Japan. It's budget in 2013 was 211 billion JPY, or $2.0 billion dollars. It has built a number of space science missions, and participated in international projects like the ISS.

Government Programs - United States

The US government is a major funding source and operator of space projects. The funding ultimately originates from the nation as a whole through taxation. Funds are appropriated annually through the US Congress to several Executive branch agencies. Activities of the US government in general are covered through by the annual **Aeronautics and Space Report of the President**, issued annually. Budget figures in that document may not include classified military programs, so we can say the government total is $30 billion or more.

NASA

The **National Aeronautics and Space Administration**, or **NASA**, is the primary civilian space agency of the US government. It was originally created in 1958, and currently is authorized under **Title 51--National and Commercial Space Programs** (Public Law 111-314, 18 Dec 2010). Funding is appropriated annually by the US Congress to operate NASA. For 2016 it is contained in **H.R. 2029** at pages 74-77. NASA is headquartered in Washington, D.C. and currently has 17,345 staff and a number of **Centers and Facilities**. A large part of its approximately $19.3 billion annual budget is spent on contracted work. Wikipedia has an extensive **List of NASA Program Articles**, and the **NASA Missions Page** lists past, current, and future missions.

Past Projects

When the United States decided to create NASA as a civilian agency in 1958, it inherited facilities and staff from the Department of Defense and the previously existing **National Advisory Committee for Aeronautics**, which dates back to 1915. The agency's **History Office** published a self-history covering 1915-1990. It was originally in paper form as NASA special publication SP-4406, then online as **Orders of Magnitude**, and also available as a **PDF file**. Numerous other history publications have been produced by this office. Care should be taken to recognize bias in any internally written history, but factual data is likely to be correct.

Current Projects

The **Current and Recent Budgets** page provides data on recent and current activities from a financial standpoint. The **2017 Complete Budget Estimates** from that page, and similar ones for past years under the "Previous Years' Budgets" section, also provide detailed descriptions of on-going projects.

NASA operates under three main mission directorates. The **Aeronautics Research Directorate** operates a number test facilities and research projects. Some of these have application to space launch. The **Science Mission Directorate** manages numerous projects and missions related to Earth science, heliophysics (the Sun and space environment), planetary science, and astrophysics. Finally the **Human Exploration and Operations Directorate** develops and operates the very well known manned programs, and less well known science and technology projects related to human exploration. The division of funding (in US dollars/year) in 2016 and major projects include:

Scientific Research ($5.59 billion)

NASA Science programs include Earth Science, Planetary Science, Astrophysics, and Heliophysics project divisions, and also satellite work NASA performs for other government agencies (NOAA and USGS). The science and technical parameters of astronomy and planetary science projects funded by NASA are listed below under not-for-profit programs for several reasons. The scientists who are the end users of the data generated are often at universities and other not-for-profit institutions, the terrestrial projects often have multiple funding sources, and programs in the sciences are better understood as a whole rather than divided by funding agency The budget and organizational details are noted in this section.

Earth Science - This division is directed at studying the Earth, but the spacecraft hardware and instruments are often similar to what is needed at other locations. In addition to the specific projects noted below, the division funds data archiving, organization, and delivery, advanced technology development, and application of the data to human needs.

- **Earth Science Research** - Scientific research aimed at understanding the Earth, its components, interactions, changes, and how these affect life. It includes a substantial effort at high performance computing for modeling and analyzing the large volume of data generated by satellites and other instruments.
- **Earth Systematic Missions** - This is a collection of Earth-observing research satellite missions which collect different types of data on the Earth as a whole. Some of the projects are jointly between NASA and other US agencies, or other national agencies. New satellites include **Surface Water and Ocean Topography** and **NASA-ISRO Synthetic Aperture Radar**. It also includes operating numerous current satellites and instruments, and scientific analysis of the collected data.
- **Earth System Science Pathfinder** - These are small to medium projects and satellites aimed at collecting Earth data, using new and emerging priorities and technology
- **Venture Class Missions** - This includes suborbital, small missions, and instruments for earth science investigations.

Planetary Science - This division is directed at studying the content, origin, evolution, and potential for life of our Solar System and planetary systems beyond our Sun. To get better data than observing from a distance (i.e. astronomy) it has a progressive strategy of flyby, orbiting, landing, roving, and sample return for each object. Generally astronomy is how original detection of new objects occurs, although the **Near Earth Object Program** within this division is directed to find close and possibly hazardous objects down to 140 meters in size. It funds general science research and technology development projects, plus specific groups of spacecraft and missions noted below.

- **Discovery Program** - This includes relatively smaller missions with shorter development times. It includes the **Dawn** spacecraft studying Ceres, **InSight** Mars lander, and **Lunar Reconnaissance Orbiter** mission.
- **New Frontiers** - This includes medium size, high priority science missions, including the **New Horizons** Pluto flyby, **Juno** Jupiter orbiter, and **OSIRIS-REx** NEO sample return missions.
- **Mars Exploration** - This program seeks to understand the past, present, and future habitability of Mars through a series of incremental missions. It includes continuing operation of the **Mars Odyssey** and **Mars Reconnaissance** orbiters, **MAVEN** atmosphere sampler, and **Opportunity** and **Curiosity** surface rovers. In development is the **Mars 2020 Rover**.
- **Outer Planets and Ocean Worlds** - This program is aimed at understanding the origin and evolution of the outer solar system. It includes the current **Cassini** mission to Saturn, development of the **Europa Mission** to study that moon of Jupiter, scientific research, and planning for future missions.
- **Technology** - This program develops improved technology for future planetary missions. It currently includes projects in electric propulsion, radioisotope generators, and mission management software.

Astrophysics - The areas of research for this division include the Universe as a whole, the evolution of galaxies, stars, and planets, and the characteristics, habitability, and presence of life on extrasolar planets. It funds general research and data analysis from NASA missions, as well as specific programs as follows:

- **Cosmic Origins** - This program investigates the evolution of the Universe and its components. Major projects include the **Hubble Space Telescope**, **SOFIA** airborne telescope, and **Spitzer** infrared telescope.
- **Physics of the Cosmos** - Investigates how the universe behaves under extreme conditions. It supports two operating missions: the **Fermi** gamma ray telescope and **Chandra** X-ray observatory, along with supporting science and technology work.
- **Exoplanet Exploration** - Searches for and tries to characterize planets outside our solar system. Much of this research happens with ground-based telescopes such as Keck and LBT, but this program includes the **Kepler** space telescope, which is a dedicated planet finding mission. Work on the **WFIRST** large space telescope is in the early stages. The program includes science and technology development for future instruments and missions.
- **Astrophysics Explorer** - This program funds smaller, dedicated science missions. Current projects include the **TESS** exoplanet survey telescope, participation in international projects, and continued operation of some previous missions.

James Webb Space Telescope - This is a 6.5 meter diameter red to infrared space telescope, the largest of its kind to date. It is currently planned to launch in October 2018. The extreme cost of this project ($8.8 billion total) has generated congressional scrutiny and separate budget tracking from the rest of the Astrophysics science budget.

Heliophysics - This division studies the Sun, the space environment between bodies, and that surrounding planets. Like the other science divisions, it funds topical science research, data analysis and archiving, and more specific programs noted below. Sub-orbital launches and continuing operation of previously launched satellites are also funded by this division. This includes the two **Voyager** spacecraft, which have completed their planetary science work and are now primarily collecting data on the space environment.

- **Living with a Star** - Studies the interaction of the Sun and Earth and its effects on life and society. Current projects include the **Solar Probe Plus** which will fly very close to the Sun, and the joint **Solar Orbiter** mission with ESA.

- **Solar Terrestrial Probes** - Studies the plasma environment between the Sun, Earth, and solar system. Projects include the four **Magnetospheric MultiScale** spacecraft, and continuing operation of other missions.
- **Heliophysics Explorers** - Includes small to medium targeted science missions. It includes the **ICON** ionospheric explorer, and operation of a number of previous missions.

Aeronautics Research ($640 million)

The NASA **Aeronautics** Research Directorate works to improve aviation and high speed flight. Entry, descent, and landing through an atmosphere from space is studied elsewhere under Space Technology. In addition to general management, innovation, and education activity, specific programs include:

- **Airspace Operations and Safety** - Performs research and development to increase the throughput, efficiency, and safety of the National Airspace System.
- **Advanced Air Vehicles** - Performs research for new generations of civil aircraft, including supersonic and hypersonic technology. Note that the US Department of Defense does similar work for military purposes.
- **Integrated Aviation Systems** - Focuses on experimental flight research using integrated systems, flight test ranges, and aircraft. This complements high performance computing for analysis and simulation, and physical testing in wind tunnels and propulsion test facilities. Testing is coordinated with the DOD, who have similar facilities and test needs.
- **Transformative Aero Concepts** - Solicits and works on multi-disciplinary and revolutionary ideas, from original concepts through small scale ground and flight testing. It also advances computational and experimental tools and technologies.

Space Technology ($686 million)

This category funds space technology development from very early concepts to flight demonstration, after which it would be incorporated into operational programs. It is managed by the **Office of Chief Technologist** (OCT). The actual work is done within NASA centers, and the commercial and not-for-profit sectors. The OCT develops roadmaps to plan their work, which were reviewed in a **2012 Report** by the independent National Academies, and guides funding plans. In addition to fostering technology transfer across NASA programs, and externally across other agencies and society in general, specific programs include:

- **Small Business and Technology Transfer Program** - These are competitively awarded contracts to small (<500 employee) business for innovative technology. They progress from initial merit and feasibility through commercialization.
- **Early Stage Portfolio** - Invests in about 400 basic and applied research and early technology development activities. This includes external research grants and innovative concepts, internal innovation at NASA centers, and Centennial Challenge prize competitions.
- **Game Changing Development** - Matures technologies from early stages to flight demonstration. It includes hardware topics like robotics, manufacturing, and materials. It also includes mission topics like entry, descent and landing; future propulsion and power, and destination systems and instruments.
- **Technology Demonstration Missions** - Demonstrate readiness for operational use using prototypes and demonstration units in the relevant space environments. Current examples include satellite servicing, deep space optical communications and atomic clock, non-toxic propellants for small thrusters, solar-electric propulsion, supersonic decelerators, composite materials, long-term cryogenics, in-space robotic manufacturing and assembly, and small spacecraft technology.

Human Exploration ($4.03 billion)

This budget account is focused on human exploration beyond low Earth orbit and transport to the existing International Space Station (ISS). The program is based on goals in the **NASA Authorization Act of 2010**.

Exploration Systems Development - About three quarters of exploration funding goes to three major programs under this heading:

- **Orion Multi Purpose Crew Vehicle** - This is a modular launch abort system, crew capsule, and service module designed for long duration deep space missions. It is currently in development by **Lockheed Martin Corporation** for NASA, and had an early flight test in 2014. and uncrewed launch on the SLS in Nov 2018.
- **Space Launch System** (SLS) - This is a heavy-lift rocket for launching payloads beyond low Earth orbit. The initial design uses used Space Shuttle program **SSME** (RS-25) engines and modified Shuttle **Solid Rocket Boosters** for the lower stage, and **RL10** second stage engines.Contractors include **Boeing**, **Aerojet Rocketdyne**, and **Orbital ATK**. First launch is planned for Nov 2018.
- **Exploration Ground Systems** - Provides funding for the SLS specific parts of the launch site at the Kennedy Space Center. This work is in parallel with SLS development.

Human Research Program - Conducts research and develops technologies that allow humans to travel safely and productively in the environment of space. It was allocated $142 million funding in 2015.

Advanced Exploration Systems - Funds high priority projects for future human missions. Areas of work include life support, deep space habitation, advanced propulsion, landing systems, and resource prospecting and processing. They were allocated $189 million funding in 2015, but this is expected to significantly increase in coming years.

Space Operations ($5.03 billion)

This part of the budget includes operating the International Space Station, the space communications network, and launch and test operations on Earth.

International Space Station (ISS) - As the name indicates, this is an international program. This program covers the NASA funded portion, which amounted to $1.52 billion in 2015. The program is expected to continue through fiscal year 2022. One domestic purpose of the ISS is to promote commercial research and transportation in Earth orbit. Tasks within the program include:

- **ISS Sytems Operations and Maintenance** - Responsible for assembling, operating, and maintaining the ISS with an onboard crew of 6. This includes mission planning, ground monitoring and communications, spares and logistics, crew training, mission integration (allocating all orbital resources), handling system failures, and safety and mission assurance.
- **ISS Research** - Funds biological and physical research which benefits from the orbital environment. Major areas of research include human microgravity effects, plant and microbiology fluid, thermal, and particle physics, combustion, and materials processing. It is also a platform for Earth and space observation, although not unique in this aspect. The ISS is also being used as a demonstration platform for robotics, life support, and fire safety technology. The US portion of the Station has been designated the **ISS National Laboratory**.

Space Transportation - This program provides current and future transportation to Earth orbit of astronauts, science experiments, supplies, maintenance hardware, propellants, and return of wastes. The current destination is the ISS. Funding was $2.25 billion in 2015.

- **ISS Crew and Cargo Transportation** - Purchases commercial transportation to the ISS from Orbital ATK, SpaceX, Sierra Nevada, Boeing, and Roscosmos. Note that other ISS partners provide their own transportation and payloads.
- **Commercial Crew Program** - Supports development of commercial US crew transport by Boeing and SpaceX. Flights are expected to start in 2017.

Space and Flight Support - This includes other operational activities, including the communications network, launch services, propulsion testing, and human space flight.

- **Space Communications and Navigation** - Operates the communications network and navigation services for all NASA spacecraft. It includes the Deep Space Network for long range communication, NASA-owned Tracking and Data Relay Satellites in synchronous orbit, and development of future laser communications for higher data rates.
- **Human Space Flight Operations** - Provides astronaut crew training, and health and safety monitoring.
- **Launch Services** - Manages the Expendable Launch Vehicle program. Actual cost of individual rockets is included under the project or mission that requires them, while contracting, and launch site operations are funded here.
- **Rocket Propulsion Test** - Operates four rocket engine test facilities.

Support Activities ($2.87 billion)

This portion of the NASA budget covers general staff and support not tied to a particular program or project. It includes public education activities, agency staff, facilities construction and maintenance, and inspector general.

Education - Interacts with all levels of education institutions, museums, and science centers to promote interest and programs in science, technology, engineering, and mathematics.

Safety, Security, and Mission Services - Operates NASA centers and headquarters, including 17,300 agency staff and IT.

Construction - Includes new construction, demolition, environmental compliance, and repair of existing buildings and other facilities.

Future Projects

The **2014 NASA Strategic Plan** provides a general description of their planned future objectives. These include:

Expand the frontiers of knowledge, capability, and opportunity in space:

- Expand human presence into the Solar System, and to the surface of Mars.
- Conduct research on the International Space Station.
- Employ U.S. commercial capabilities to deliver cargo and crew to space.
- Understand the Sun and its interactions with Earth and the Solar System.
- Ascertain the content, origin, and evolution of the Solar System, and the potential for life elsewhere.
- Discover how the universe works, explore how it began and evolved, and search for life on planets around other stars.

- Mature crosscutting and innovative space technologies.

Understand the Earth, and develop technologies to improve life on our home planet

- Advance aeronautics research for safe and sustainable aviation.
- Advance knowledge of Earth as a system.
- Optimize technology investments, open innovation, and infuse technology for national benefit.
- Advance STEM education and workforce pipeline.

Manage the agency's people, technical capabilities, and infrastructure.

This includes attracting a skilled workforce, innovative work environment, and necessary facilities, tools, IT, services, and capabilities.It also includes maintaining safety

A general set of objectives is not the same as what will actually or likely get funded and built. Better detail for the next five years can be found in annual budget requests, such as the one for **Fiscal Year 2017**. The duration of current and upcoming projects often runs longer than the 5 year budgeting horizon. So individual agency offices and projects usually have plans and schedules which run somewhat longer. NASA's funding is on an annual basis, and therefore unpredictable to some degree. The nature of research and technology development, which involves finding out new things, is also unpredictable. Therefore detailed long range planning does not make sense. Instead, long range goals and directions are set, with detailed plans made approximately 5 years ahead. Additionally, concrete plans for future projects would immediately raise funding issues with the US Congress, who would have to approve the money for those projects.

Continued on page 2 →

Section 1.9 - Existing Programs (page 2)

← Back to Page 1

Government Programs - United States (continued)

Department of Defense (DOD)

The space projects of the US Department of Defense (DOD), are primarily managed by the United States Air Force (USAF) **Space Command**, with some assistance from other branches and agencies within the department. It is difficult to get a full view of DOD space-related projects, since it only makes up part of the department's total activity. The **2016 DOD Budget Overview** reports $7.1 billion for space investment (page 5-3). The Space Foundation **2015 DoD Space Budget Comparison** indicates a total of $9.75 billion, with detailed tables. However, much of the space-related activity is classified, and not included in these totals. Between 2012 and 2013, reported DoD space spending dropped from $27.5 to 10.8 billion, and this difference may represent the classified amount.

What is known publicly is the DOD procures and operates a number of satellites and supporting launch systems in the areas of communications, weather, nuclear detection, mapping, and navigation. It also funds a significant amount of scientific research and engineering development either directly or indirectly related to space projects. Examples of indirect efforts include high-speed air-breathing propulsion and radiation-hardened electronics.

Past Projects

Ballistic missiles were developed for military purposes rather than space launch. The requirement to deliver a bomb thousands of miles requires reaching about 80-90% of Earth orbit velocity. Therefore ballistic missiles could be adapted for space launch, usually by adding or enlarging the upper stage. A number of them were adapted this way in the late 1950's to 1960's. As the size of space satellites grew far beyond that of nuclear bombs, those launchers were repeatedly modified, to the point that some of them no longer retained any original parts except the name. Besides being used for delivery of DOD spacecraft, the same launch sites and vehicles have often been used for non-military launches. Ballistic missiles are not designed to be used more than once, and their value is relative to the target they intend to destroy. So an unfortunate side effect of using modified ballistic missiles for space launch is they were not optimized for cost. In fact, expensive throw-away hardware is the exact opposite of optimized for cost. Since the previous technology and experience was gained in that environment, later space projects have had to struggle to overcome this history.

National Science Foundation (NSF)

The **National Science Foundation**, or NSF, is an independent federal agency which funds research in all fields of science and engineering, except medical sciences, which have their own agency. The expected funding from the **2016 Budget Request** totals $7.72 billion, a $380 million increase over 2015. The space-related part of their budget was about $460 million in 2014. Although all knowledge is a seamless whole, and much of it applies to space projects, certain fields are currently more directly related than others. The NSF offices in this category include:

Engineering Directorate (ENG)

Funds research in all the fields of engineering. Much of this can apply to space projects because most fields of engineering are used in such projects. Particular areas of interest are:

- Chemical, Bioengineering, Environmental, and Transport Systems (CBET), which funds research in chemical, mechanical, and aerospace topics.
- Civil, Mechanical, and Manufacturing Innovation (CMMI), which funds topics named it it's title plus materials design.
- Electrical, Communications, and Cyber Systems (ECCS), which funds electronics, communications, power sources networking, and robotics.

Mathematical and Physical Sciences Directorate (MPS}

Funds the scientific fields of astronomy, chemistry, materials research, mathematics, and physics. This includes the **Astronomical Sciences Division** (ASD), which funds research in astronomy, including planetary science, and contributes funding or operates a number of ground-based observatories such as ALMA and LSST. Details of astronomy projects are gathered under the not-for-profit section, because typically they have multiple funding sources for the instruments and researchers.

Department of Commerce - NOAA

The US Department of Commerce's National Oceanic and Atmospheric Administration (NOAA) operates weather and earth observation satellites via its **Satellite and Information Service**. Weather satellites are located both in low polar and geostationary orbits, which supply different levels of resolution and temporal coverage. NOAA also operates the JASON sea-level measurement and DSCOVR solar wind satellites. Space related funding was about $2.1 billion in 2014.

Department of Energy (DOE)

The US **Department of Energy** is a cabinet level department formed in 1977 to consolidate nuclear and other energy activities of the United States. It performs research into all forms of energy and some scientific research. Almost all space projects require energy to function, so some of this research is relevant. Details of department's current programs and funding can be found in it's **2017 Budget Request**. Space-related parts of their activities include:

Energy Efficiency and Renewable Energy

This office performs research into all forms of energy supply and energy efficiency. Total funding in 2016 was $2.07 billion. Programs of particular interest include:

- **Solar Energy Techologies** - Develops photovoltaic and concentrated solar power, aiming to reduce cost and increase efficiency. Given the high solar flux in the inner Solar System, many space projects use this as their main power source. Note that by 2016 ground solar energy has developed into a large worldwide industry, and much research and development now occurs outside US government funding.
- **Advanced Manufacturing** - Participates in the **National Network for Manufacturing Innovation**. Mining and manufacturing will be increasingly important in future space projects, because of the high leverage of using local resources rather than launching everything from Earth.

Nuclear Energy

This office performs research into Nuclear energy sources in particular. Total funding in 2016 was $986 million. Programs of particular interest include:

- **Small Modular Reactors** - Since mass is a concern for space projects, power sources smaller than terrestrial nuclear power plants are desirable.
- **Radiological Facilities Management** - Provides Radioisotope Thermal Generators, which are used as power supplies for space missions. They are developed, built and tested at the **Idaho National Laboratory** and several other DOE national laboratories.

Science

This office funds scientific research and scientific tools to understand nature and advance energy security. Total funding in 2016 was $5.35 billion. Much of the work occurs through a network of National Laboratories. Programs of particular interest include:

- **Basic Energy Sciences** - Includes research in materials science, engineering, chemical, geoscience, bioscience, and large scientific user facilities.
- **Fusion Energy Sciences** - Supports research to develop fusion as an energy source. Although the Sun is a natural fusion reactor, solar energy is not available deep underground, or at large distances from the Sun, so artificial fusion would be useful in these locations.
- **Nuclear Physics** - Supports research into all types of nuclear matter. This includes artificial and purified isotopes, which can have space applications.

ARPA-E

This agency within the Department of Energy focuses on early-stage energy technologies which have high potential and high impact, but are too early for private investment. Total funding in 2016 is $291 million. Nearly all space projects require energy to operate, so improvements in this technology would be useful.

Commercial Programs

Commercial space projects are part of the more general "Aerospace and Defense" business sector. An **Annual Overview** of the entire sector is compiled by Price Waterhouse Cooper. There is considerable overlap between items used in space projects and those developed for other parts of the sector, and often the same company does both types of work. The sector as a whole had worldwide revenue of US$ 729 billion in 2014. Note this excludes undisclosed classified or private projects, and that supplier sales which end up in final products in the industry are double-counted.

The **Satellite Industry Association** is a US-based trade association for commercial space businesses. It has a **2016 Industry Report** which indicates the global space industry in 2015 was US$ 335 billion, of which the satellite portion was US$ 203 billion. The latter represents 9.2% of global telecommunications revenue. The vast majority of satellite revenue is from services and ground equipment, such as for satellite television. Satellite hardware manufacturing and launch accounts for US$ 21.8 billion of the total. Since 208 satellites were launched in 2014, that implies the average unit cost to build and launch is US$ 105 million. For comparison, the average prices of Boeing passenger aircraft range from US$ 60-330 million.

By far the largest segment of commercial programs is for communication services, of which consumer satellite TV is the largest single component at 2014 revenue of US$ 95 billion. All other satellite services amounted to US$ 27.9 billion. This included satellite radio and broadband, commercial communications, mobile voice and data, and remote sensing. Unclassified satellite manufacturing averaged about US$ 13.75 billion per year in recent years, and launch revenue averaged about US$ 5.2 billion. Both are counted in year of delivery, while actual costs usually are distributed over several years. Ground equipment to receive from or transmit to satellites accounted for US$ 58 billion in 2014.

Future commercial programs which are in research and development but not yet significant in revenue include space tourism, orbital mining, and energy transmission.

Not-for-Profit Programs

The not-for-profit sector includes activities performed for the general benefit of humanity, most often at universities and research foundations. At present, most such space programs are related to **Astronomy**, and the cross-discipline field of **Planetary Science**. Astronomy is partly related to space systems because that is the science which studies the Universe i.e. all of space outside the Earth. It is also partly because the Earth's atmosphere, gravity, and day/night cycle interfere with certain types of instruments and observations, so they must be performed in space themselves. Individual researchers can work on their own, but the larger projects, such as new telescopes, are often a mixture of private and government funding.

Planetary science is the study of objects and systems which orbit stars. Originally it was purely a subset of astronomy and restricted to the Solar System, as no other planetary systems were known. As better instruments and close-up observations have been made, we have progressed from merely determining orbit and approximate size of planets and moons to detailed mapping and geology. So planetary science now draws heavily on the **Earth Sciences** to understand the history and development of these bodies. In recent decades **Circumstellar Discs** and **Exoplanets** have been detected around other stars, extending the study beyond our Solar System. More recently, the first few **Rogue Planets** have been discovered. These objects are too small to be stars, but not attached to any stellar system. They were either ejected from star systems or originally fomed as unattached objects.

Astronomy Programs

The US National Academies compile a **Decadal Survey** every 10 years, laying out priorities for astronomy and astrophysics. This is a good starting point to review current and near-term programs in this field. Other science and engineering departments at universities do research related to space, and there are some smaller foundations dedicated to space research. Although much of the funding for astronomy comes from government sources, we list all of the programs here to give a better view of the field as a whole.

Astronomy programs can be roughly sorted by location and wavelength of the instruments. Locations include ground, airborne, and space. The latter two are more expensive, but are used to get above interference from the Earth's atmosphere. The **Electromagnetic Spectrum** ranges from long radio to very short gamma, and instruments exist to cover most or all of it. The location of the instruments, which are grouped into **Observatories**, is nowadays where the viewing is best for that device. This is usually different than where the funding organizations or the astronomers who make use of the data are.

There are a great number of observatories in use, since even small privately owned ones can collect useful data. Wikipedia has a lists of **Astronomical Observatories** and **Space Observatories**, the latter meaning located in space (all of them look at space). A few of the more significant ones which are relevant to future space programs are noted below. These are mainly ones that look at our own and nearby planetary systems.

Current Projects

Hubble Space Telescope - This most famous of telescopes is a 2.4 meter UV to infrared space telescope launched in 1990 and expected to operate to until either equipment failure or orbit decay terminates the mission, likely in the 2018-2024 period. In addition to other science, Hubble has been used to find and examine outer solar system objects, extrasolar planets, and protoplanetary disks. Hubble is mainly funded by NASA and ESA, and scientific operations are managed by the **Space Telescope Science Institute**.

Spitzer Space Telescope - This is a 0.85 meter infrared space telescope launched in 2003. Its original helium supply, used to cool the instruments, ran out in 2009, so at present only the shortest wavelength instruments are still operating. At some point equipment failure will end the mission. Spitzer observed many solar system and extrasolar objects. It is funded by NASA.

Kepler Mission - This is a 0.95 meter visible light space telescope operating since 2009 through at least 2016. It is designed to detect planets orbiting other stars which cross in front of the star (transit) and dim the star's light. Since that only happens when the orbits are edge-on to us, Kepler can only detect a fraction of the total planets in the direction it is looking. The total number of planets can be estimated from the fraction it can see. Kepler is funded and operated by NASA.

Stratospheric Observatory for Infrared Astronomy (SOFIA) - This is a 2.5 meter infrared airborne telescope mounted on a 747 aircraft which first saw operation in 2010. At its operating altitude of 13.7 km it is above most of the absorption caused by the Earth's atmosphere. It will study, among other things, the formation of stars and planets, the interstellar medium, and planets and small objects within our own solar system. It is funded 80% by NASA and 20% by the **German Aerospace Center** (DLR).

Atacama Large Millimeter Array (ALMA) - This is an array of 66 x 12 and 7 meter radio telescopes that function as an interferometer to combine their signals and act as a large single instrument up to 14 km across. It began scientific observations in 2011 when partially built, and was fully operational in March 2013. It is a general purpose radio telescope in the 0.35-10 mm wavelength bands.

Among the types of observations it is making are circumstellar dust and planetary systems around other stars. ALMA is jointly funded by the US, Europe, Japan and is hosted by Chile. The **ALMA Website** has additional information about the project.

Gaia Observatory - This is a space satellite with instruments to measure the accurate position, motion, brightness, and spectra of about a billion stars. It was launched in late 2013. It is likely to find many planets by the wobble they cause in their parent stars. It is also expected to find and accurately measure the orbits of many small objects in our Solar System. The ESA sponsors this mission, and their **Gaia Website** has additional details about it.

Radar Astronomy - A few instruments are used to actively send out radar signals and measure the return. The timing of the return signal provides extremely accurate location and detailed shape information. The latter comes from timing differences from parts of the object that are at different distances, and repeating the measurements as the object rotates. Signal intensity falls as the 4th power of distance, so this technique has been limited to approximately 500 nearby objects, mostly Near Earth Asteroids.

Future Projects

James Webb Space Telescope (JWST) - This is a 6.5 meter visible to mid-infrared space telescope expected to launch in 2018. It has a number of science objectives, including observing extrasolar planets, brown dwarfs, and outer Solar System objects among the more relevant ones for space programs. The instruments have a fairly narrow field of view, 2x4 minutes of arc or about 1% of the area of the Moon as seen from Earth. So it will do targeted observations rather than wide surveys of the whole sky.

Wide-Field Infrared Survey Telescope (WFIRST) - This is a 2.4 meter near-infrared space telescope proposed for launch in the mid-2020's. The Science Definition Team membership comes from many universities and several independent observatories and US government centers. They released a **Final Report** in February 2015 describing the science goals and telescope design. Among the science goals is to perform a search for planetary systems in our Galaxy using gravitational microlensing. The program would be primarily funded by NASA, but the data would be used by astronomers worldwide.

Thirty Meter Telescope (TMT) - This is a 30 meter near-UV to mid-infrared (0.31 to 28 μm) ground telescope with an estimated completion after 2022. It will be made of nearly 500 1.4 meter hexagonal mirror segments on a single large mounting to be built at Mauna Kea Observatory in Hawaii. It has multiple science goals, including extra-solar planet and Kuiper Belt measurements. It is jointly funded by multiple foundations, universities, and national governments. The project is currently (2016) delayed by local protests and the construction permit process.

Giant Magellan Telescope (GMT) - This is a 24.5 meter equivalent visible and near infrared ground telescope with an estimated operational date of 2021. It is made up of seven 8.4 meter mirrors on a single large mounting, because that is the largest mirror size that can currently be made in one piece. As of 2016 one of the mirrors has been completed and three others have been cast. Construction the observatory was started in Chile in late 2015. Like other large telescopes it has multiple science goals, including imaging of planets around nearby stars. It is currently funded by a group of universities.

Large Synoptic Survey Telescope (LSST) - This is an 8.4 meter visible and near-infrared ground telescope with a 3200 megapixel camera which has begun construction. The **LSST Construction Schedule** shows a completion date of early 2022. It is a general purpose survey telescope which the **LSST Science Book** estimated could find 90% of Near Earth Objects larger than 140 meters in 12 years (ie 2034) if 15% of it's time is used for that purpose (section 5.11.1).

European Extremely Large Telescope (ELT) - This is a 39 meter visible and near-infrared segmented ground telescope expected to be operational about 2024. This is the largest optical telescope currently planned. Its science goals include finding extrasolar planets. It is being built by the **European Southern Observatory**, which is funded by 16 nations and one host country (Chile), where the telescope will be located.

Planetary Science Programs

Approximately eighty **Graduate Schools** around the world offer degrees and perform research in planetary science. Some of them participate in government-funded planetary and space observatory missions by building instruments or supplying scientists. Others participate in or run ground-based observatories, perform independent theoretical research, or analyze data generated by the observatories and missions.

Section 1.10 - Future Projects

Section (1.9) described existing space programs and projects. This section lists categories of space activities and mission objectives for future programs and projects. Parts 2 and 3 of the book will describe the transport and engineering elements to implement these activities. They also discuss how to design, build, and operate complete systems that perform the desired functions. A given project, location, or mission may include more than one of the items below. The multiple activities can be present but not otherwise interlinked, in the same way a large office building may house multiple unrelated tenants. In that case the project may share joint services like electrical power among the various activities. Alternately, the multiple activities may link to each other in parallel, such as retrieving asteroid materials and refueling the tug from those same materials. The activities may also be connected in a logical sequence, where one activity follows another, such as building new exploration or mining vehicles from previously produced inventory materials. We call any program with multiple different activities a **Complex Program**, and an extensive example is provided in Part 4. Many other such complex programs are possible, but our example is intended to be both a tutorial and a reasonable proposal for future activities.

The activities listed below are somewhat in order of idealized complexity. An actual program that implements multiple items would likely choose a different order which makes sense from a design and schedule standpoint. As civilization expands into space, eventually we will incorporate all the industries and activities that occur on Earth today. So this list is not comprehensive. Instead, it includes the more significant items that are particularly space related.

Bulk Mass

Bulk mass is matter which is undifferentiated into manufactured components, either in the raw unprocessed state, or processed into a refined product. The quantity of bulk matter is a main variable, as all the bulk material of a given type is equivalent. The Earth has a deep gravity well, which requires 62.5 MJ/kg to escape. This is currently difficult and expensive. Local sources of bulk materials, that are already in space near to where you need them, can often be delivered and processed for less total energy. Solar energy is abundant in the inner Solar System in quantity, and available nearly everywher. So the delivery and processing in theory can be low cost. So there is an economic incentive to obtaining bulk materials in space. Bulk materials can be used for propellants and fuels, serve as radiation shielding, or treated as mineral ores for extraction and further processing into useful products.

NASA has termed use of local materials in space **In-Situ Resource Utilization** (ISRU). They have set up an **ISRU Project Office** and are doing some early research on methods. The ISRU name is more palatable to some members of committees who decide NASA's budget than "space mining", which sounds too much like science fiction. Despite that, Title IV of Public Law 114-90, also known as the **Space Resource Exploration and Utilization Act of 2015** makes provisions for private commercial recovery and ownership of space resources. Asteroids have negligible gravity wells, and by early 2016 over 14,000 of them in the **Near Earth Asteroid** (NEA) category have been discovered. Different asteroids also have a variety of useful chemical compositions. This combination has led to the most interest in **Asteroid Mining** relative to other space locations, but any location with available materials and energy is a candidate. NASA is also in the early stages of developing an **Asteroid Redirect Mission** to bring back a boulder from an NEA to a safe lunar orbit, where it can later be examined in detail.

Propellants and Fuels

Propellants are the materials consumed and ejected in many types of propulsion systems. They often make up a large or dominant portion of space system mass. Propellant applications include space transport, ballistic transport, aerodynamic transport, surface transport, and portable power. Due to the wide range of applications, environments for their use, and sources of raw materials, there is also a wide range of potential propellants. Specific propellants for space transport are discussed in Part 2 of this book. The other applications are listed below.

Fuels are energy sources reacted with materials in the environment, such as air-breathing engines, or with an oxidizer in the case of chemical rockets. Nuclear fuels develop energy by internal reactions, and do not require another material to react with. Of particular interest for advanced uses are fission fuels such as Uranium or Thorium, and if fusion power is developed, light isotopes for those reactions. This is the energy released by nuclear reactions is very high relative to chemical ones.

Production of propellants and fuels from local resources is called **In-Situ Propellant Production**. To date, all propellants and fuels have been supplied from Earth. Using local resources will often prove less expensive. The mass overhead for delivery is reduced by using depots at intervals along a trip. If all the propellant mass is carried from the start, more of it has to undergo a velocity change. So you have the overhead of using some of it up to move the remainder. By refilling at intervals, the on-board mass is less, and so the mission overhead of moving it is less. The supply sources for the propellants are close to or at the depot locations, and delivery is by high-efficiency electric propulsion. Therefore the supply overhead is also low.

Besides conventional space transport, other propellant and fuel applications include:

- **Ballistic Transport** - On an undeveloped or heavily cratered terrain, attempting to drive across the surface will be difficult, and for long distances will be slow. Ballistic transport uses similar methods as space transport, but instead of going to and from orbit, or between orbits, is used in sub-orbital trajectories to go from one point to another. The lower velocities potentially allow for simpler propulsion or higher payloads. This method is especially useful on smaller bodies where the velocity requirements are lower

- **Aerodynamic Transport** - On bodies with a suficiently dense atmosphere, you can potentially use aerodynamic lift or buoyancy and a propulsion engine for transport, similar to how we use aircraft on Earth.

- **Surface Transport** - Solar power is long-lived, and can generate thousands of times the energy over it's life than a chemical fuel. However it does not have a very high power to area ratio, making it cumbersome for faster surface vehicles, and batteries can take a significant time to recharge. As an alternative, a fuel cell or combustion engine can provide more power in a compact device and be more suitable for such vehicles. Stationary fuel stations can refuel a vehicle quickly when distributed within the vehicle's operating range.

- **Portable Power** - Solar power and/or batteries are suitable for small portable devices which do not need high powe levels. Higher power levels for short periods of time may be satisfied by propellant or fuel sources.

Radiation Shielding

Varying levels of radiation occur throughout the Solar System, including on Earth. Many locations in space have levels that are hazardous to people, or can affect electronics and materials. Natural space radiation comes from four main sources: cosmic rays, solar wind and plasma events, trapped particles in magnetic fields, and concentrations of radioactive elements. Human-caused radiation can come from nuclear and other high energy devices. The effectiveness of shielding varies by composition, with light elements protecting better for some types of radiation, and heavy elements for others. Where mass is important, optimized and processed materials may be preferred for shielding, but unprocessed bulk material is easier to supply.

Habitats, vehicles, and equipment can reduce radiation levels by their own mass, before adding additional shielding. For example, large habitats with thick walls, agricultural soil, water, and atmosphere may provide sufficient shielding as is. For vehicles or lightweight habitats, additional shielding can be added bulk propellants, water, and food supplies. The surface or subsurface of large bodies usually has sufficient local material to provide radiation shielding. The approaches to using it are building underground, or surface construction by moving enough material around and over radiation-sensitive areas. A thickness of 25 cm or more of loosely packed unprocessed material can reduce exposure to reasonable levels for humans (see **Miller et al, 2008**). In open space, where bulk materials are being processed, the unprocessed raw material, stockpiles of processed materials, and wastes after processing (slag) can all be used for shielding with proper arrangement around habitats and equipment.

Ore Delivery

An **Ore** is any natural material containing enough of a desired product to be economic to extract. On Earth, crude oil, iron ore, and crushed stone are all ores used to make other products, and by mass are the largest volume transported. In space, bulk ores and their products are also likely to be a major transport item by mass. Depending on economics and technology, bulk ore can be transported in it's raw state, concentrated in the desired components, called **Beneficiation**, and then transport the concentrate, or processed in place to a final material which is then delivered to where it is needed. Final destinations can be anywhere the ore or its products are needed. Some selected examples include:

- **Delivery to Earth** - Nearly all people and economic activity is on Earth, and that will continue to be true in coming decades. A large and economically developed population uses a lot of material resources. Some materials may become scarce enough that it is economically feasible to obtain them from space. Returning high-value materials like Platinum-group elements does not pose a significant transportation challenge, so we will look at the other end of the cost scale.

Figure 1.10-1 - 60 ton Iron-Nickel Hoba meteorite in Namibia.

Assume you want to import large amounts of Iron to the Earth, since high-grade iron ores are in finite supply, and M-type asteroids can supply nearly pure iron-nickel-cobalt alloy already in metal form. The simplest method is to aim pieces from metallic asteroids at a selected spot, and just collect the bits that make it to the ground. This requires no processing. Figure 1.10-1 shows a natural example of this. You can find other examples of surviving pieces in museum asteroid collections. Non-metal slag and volatiles will tend to burn off during re-entry. Choose a size, likely around 10-50 tons, for the pieces so that re-entry drag will slow them down, and you don't get a big crater. There are plenty of places on Earth with few people and decent access for shipping. The market for steel is about 1.5 billion tons/year. The challenge is delivering it to Earth for around $1/kg or less. You likely will need to redirect and chop up a megaton (60 meter) metallic asteroid or more, but the yield is worth at least $1 billion (1 million tons steel @ $1000/ton), which may be enough to cover operating costs for an efficient operation.

- **Delivery to Space Locations** - Because of the current high cost of launch to orbit, material in a desired space location is worth much more than most materials on Earth. It needs to be moved from where it naturally occurs to where it is needed. The Earth's gravity well requires a lot of energy to climb out of, and to date launch systems use inefficient rockets to do it with. Materials which are already nearby in space can use efficient electric thrusters for transport. On the surface of bodies, some materials can be obtained locally. Transport to and from bodies smaller than Earth can use mechanical and electric methods. These all are potentially much less expensive solutions once they are set up.

- **Oort/Rogue Object Delivery** - In the distant future, a large interstellar mission may require a lot of propellants and other supplies, due to the high velocity and long duration of the trip. In this concept, several comets or unbound rogue objects, or parts thereof, are intercepted by a propulsion unit that comes from the main vehicle, or is sent ahead from the launch point. The propulsion unit consumes part of the mass to bring the rest of the mass up to the speed of the main vehicle. The delivered mass is used to further accelerate the main vehicle and resupply other materials. This allows somewhat better velocities than starting with all the fuel at the start of the mission, since the main vehicle has less mass to accelerate. For this to work, you need to know where the objects are ahead of time, or trust that their density is sufficient to find them along your path as needed.

Industrial Capacity

Simple bulk materials like propellants and radiation shielding are useful, but don't satisfy all future project goals. **Space Manufacturing** is the production of more refined or complex goods in space, from raw materials and available energy. It is distinguished from space industry on Earth, where complex goods like satellites and rockets are made, then delivered to space ready to use. Historically, a key feature of all manufacturing is using tools to make more tools. When applied to space manufacturing, a large industrial capacity can be grown from a smaller and simpler starter set. This avoids the cost of sending entire factories into space. A simple bootstrapping example is making pressure vessels from metallic asteroid material, which are then used in chemical processing of ores.

Some production will be for **Final Goods** in the economic sense. These are end products not used in further production. An example is an orbiting greenhouse that produces food for people in space. The remainder of production is either **Capital Goods**, lasting items used in further production, or **Intermediate Goods**, which are partly finished items between raw materials and final products. When the industrial capacity is intended for more than final products, and some of the capital goods are made in space, rather than delivered from Earth, then the production and growth sequence must be optimized for factors like design cost, growth schedule, and initial mass launched from Earth.

Seed Factories

A "seed factory" is a starter set of equipment which is intended to grow to a mature industrial capacity. This is by analogy to a plant seed, which grows to a mature plant. Some portion of the factory output is directed at self-growth, and the remainder to desired final products to be used. An early NASA-funded study of the concept is reported in: Freitas and Gilbreath, eds. **Advanced Automation for Space Missions**, NASA Conference Publication 2255, 1982. Computer, automation, and communications technology was not good enough in 1980 for the intended use on the Moon, so NASA did not pursue the idea further. The concept was limited to "replication", making exact copies of the starter set until enough total capacity was reached. The study also assumed the seed factory made 100% of its own parts, was 100% automated, and could only be used in space.

The current seed factory concept includes two other methods of growth: "diversification", which is making new items not in the original starter set, and "scaling", making items of different sizes than what you start with. The assumptions of 100% self-production, 100% automation, and only used in space are removed. Therefore a seed factory may start out making only a percentage of its own items, with the remainder supplied from elsewhere, and use people directly or by remote control to do some tasks. This greatly simplifies the design of the starter set. Over time, the growing production capacity can make more of its own items, and need less supplied from outside. The idea of self-expanding production using local energy and materials applies everywhere, not just in space. It is a complex subject, and this wikibook is about space systems, so a **separate book** was started for it.

Self-expanding production can provide a large amount of leverage in terms of mass launched from Earth to final products and missions you can carry out. We therefore consider it an important concept, and will reference it extensively. Because space locations are not uniform in energy or material resources, the seed equipment and what it makes will likely be distributed in a trade network where components are optimized for location. The network will include some parts and materials that still are supplied from Earth. The various parts of the network may be owned and operated by different entities, leading to a self-sustaining space economy.

Manufactured Items

Once you have set up an industrial capacity, you want to use it to make useful products. There are as many possible products to make in space as there are on Earth. Which ones make sense to make at a given location depends on mass, complexity of production, and economic value. The following items have been suggested for relative ease of production and significant mass savings:

- **Structural Materials**- A variety of structural materials can be made from local materials in space, reducing the amount of material that has to be brought from EarthAn example is Iron-Nickel shapes like columns or plates, made from metallic type asteroids. Another is cast or sintered rock, made from Lunar or Mars surface material, or stony-type asteroids. They would be melted with solar or microwave furnaces. A third example are high strength Carbon or Basalt fibers from respective local sources.

- **Solar Sails from Metallic Asteroids** - This is a combination of structural material and transport method. To recover large amounts of material from inner Solar System asteroids, Iron-Nickel alloy found in the metallic type can be rolled into foil, and then used to make solar sails. If what you want is the metal, then it sails itself to where you want it. If you want some other asteroid material, larger amounts of sail area can be used for a cargo tug. To make the sails, you need the functions of a rolling mill - a way to heat the material and a way to force it between two rollers to make thin sheets.

> Drawbacks to Iron-Nickel sails are their higher mass compared to light alloys like aluminum-magnesium, and their reflectivity is lower in the natural state. Solar sails are also somewhat limited in the directions they can apply thrust. Advantages are the raw material is readily available in large quantities from the asteroids themselves, and it does not need a lot of processing to make into a usable form. An Aluminum or Aluminum-Silicon alloy coating can be added to increase reflectivity if desired.

- **Glass** - Besides the obvious use as windows for greenhouses, glass can be used for fiber optic cable, and for inert reaction vessels, including those which concentrated sunlight is sent through. The chemically simplest glass is **Fused Quartz**, which is pure Silicon Dioxide (SiO_2). Silicon and Oxygen make up over 60% of Lunar soil, and about 50% of stony-type asteroids, so the components are very common. However they are usually bound in silicate minerals with other elements, and require chemical or thermal separation.

- **Brick and Concrete** - Brick and concrete can be used for lower-strength construction, such as radiation shielding, thermal shelters, and landing pads, when air leakage is not an issue. Conventional brick is made by heating a mixture of sand and clay until the particles partly melt and bond together, a process called **Sintering**. Building elements can be made the same way in space, provided a sufficient source of heat and right type of ingredients can be found. Its chief advantage is simplicity. Concrete is a class of artificial stone made from varying size crushed stone, called **Aggregates**, and a binder material to hold them together. On Earth the most common binder is **Portland Cement**, a mixture of shale and limestone heated to high temperature and then ground to a fine powder. Many other binders are possible, and some of them would be useful in space. The usefulness of concrete is based on it's relatively low cost, and the ability to be cast at room temperature in a variety of shapes which then harden.

- **Chemical Products** - This includes plastics, chemical reagents, lubricants, and many other items made by chemical processing. Lunar rocks are high in metal oxides, such as Silicon, Iron, Aluminum, Magnesium, and sometimes Titanium. These metals are useful for structures, solar cells, and electronics. Converting the oxides to metals and separating the elements is called **Extractive Metallurgy**. Processes include physical ones like crushing and grinding to separate mineral grains, and magnetic separation. They also include chemical ones like liquid solutions, thermal ones using high temperatures, and electrical ones like electrolysis. The chemical solutions and electrolysis require suitable reagents. These include compounds with non-metallic elements like Calcium, Potassium, Sodium, Phosphorous, and Sulfur. These elements can be found in some types of asteroids. Plastics and lubricants are typically Carbon compounds, which are also found in asteroids.

- **Biological Products** - This of course includes food, but also non-food items like wood, and chemical outputs of micro-organisms. The oldest example of the latter is alcohol from yeast, but modern biotechnology can produce a wide variety of items. Growing food typically requires water, Carbon Dioxide, fertilizers with Nitrogen, Phosphorus, and Potassium, and trace elements. Some asteroids have water or hydrated minerals. The Moon is deficient in these, because it formed in a molten state and has a low escape velocity, and they were baked out and lost. Mars and Venus have atmospheres with high percentages of Carbon Dioxide, and Mars has Nitrogen. Asteroids are a source for the elements in fertilizers, and the trace elements.

Energy

Uses for energy in space are as ubiquitous as they are on Earth. **Solar Panels** have bee used on satellites from the earliest days, since they are modular, light-weight, reliable, and produce more energy over their life than batteries or fuel cells. Power levels have varied from a few Watts to approximately 100 KW on the Space Station. Future energy needs include larger amounts for propulsion and to operate industrial systems. Habitats, communications, and scientific equipment can be large energy consumers, and finally a major future use is exporting energy to Earth or other locations in space. Types of future energy production include:

- **Solar-Electric** - This includes existing photovoltaic solar panels, and solar thermal power systems. The latter concentrate sunlight onto a heat engine/generator combination. Solar flux is adequate in the inner Solar System, but large lightweight reflectors may be used in the outer Solar System to improve power density

- **Solar-Thermal** - Some future uses require heating rather than electricity, such as an industrial casting furnace. Concentrating reflectors can achieve this rather easily. Thermal storage using local bulk mass can bridge night-time power needs. The material is heated during the day, and the stored heat used to operate a generator at night. This avoids needing large batteries for higher power demands.

- **Nuclear Sources** - Radioisotope decay heat coupled to thermoelectric generators has been used on a small scale for scientific missions to the outer Solar System, or when daytime sunlight and batteries are insufficient on a planetary surface. A number of small reactors using thermoelectric or thermionic generators have been flown. For future applications like higher-power propulsion or surface habitats, reactors with heat-engine generators, which are more efficient, have been proposed. If fusion rectors are developed, they would be very useful in space, since hydrogen is widely available as you get farther from the Sun.

- **Beamed Power** - Civilization on Earth has a large and growing demand for energy, but fossil fuels are unsustainable. Ground-based solar panels have become popular in recent years, but orbital locations can provide seven times as much energy per panel on average. This is due to night, weather, and atmospheric absorption on the ground. A large **Space-Based Solar Power** plant can send power to the ground using an efficient microwave beam. Advantages of orbital solar power are nearly 100% operating time, and lack of Carbon emissions or nuclear risks. A disadvantage is the size of the collector on the ground is governed by the transmission wavelength and distance of the orbital station, so there is a minimum size for it to function efficiently. This can be counteracted to some extent by using shorter wavelengths or lower orbits. To be feasible for Earth, the total system (orbital and ground collector) needs to be less than 7 times as expensive as solar panels on Earth, otherwise using terrestrial panels would be less expensive. Alternate uses would be to beam power to a Lunar surface base from orbit to supplement nighttime power

    At current launch costs, it makes economic sense to beam power *up* to space by swapping the transmitter and collector locations, as power in orbit is worth more than power on the ground. In the form of visible light or microwaves this would supplement on-board power obtained from sunlight. For orbital tugs at low altitude, the supplement is especially useful as the Earth's shadow covers 40% of typical low orbits.

    Laser power transmission is a future possibility. Beam generation is less efficient, but it can be focused more easily over long distances due to the shorter wavelength. Uses range from powering launch vehicles from the ground, to interstellar missions using the Sun as a gravitational lens. Given a suitable atmosphere, for example Carbon Dioxide rich ones like Venus and Mars, the atmosphere might be used as a lasing medium to generate powerful beams. This concept has not been explored in detail as far as is known.

Engineered Environments

All of space, and many parts of the Earth, have conditions not suited to humans, or higher life in general. On Earth we apply the various fields of engineering and technology to modify the natural environment in specific locations, such as buildings, ships, and aircraft. Common modified conditions include temperature, protection from weather, and pressure (in the case of aircraft and submarines). In space, we must modify additional conditions like atmospheric composition (or entire lack of atmosphere), gravity level for long term stays, radiation levels, and other parameters. In some circumstances the environment would be set up for plants (high CO2 ratio), or machines, rather than humans. To date, engineered space environments have involved short-term travel in vehicles, even shorter times in space suits, and up to a year and up to 6 or so people in space stations.

Future projects may include larger populations and longer stay times, construction of habitats in space rather than launching already-built units, and production and recycling of basic needs like air, food, and water. Reasons include longer-duration exploration and science, commercial and industrial activities, and the desire to live in interesting and unique locations. Future space environment projects include:

- **Space Habitats** - Humans evolved on Earth, so it is the only place we know of where we can survive, even for a short time, without the help of technology. Parts of the Earth, such as much of Antarctica, are lethal even in a short time without at least clothing. A **Habitat** in nature has the right conditions for particular species to live. Artificial habitats, such as homes and greenhouses, are purposely built artifacts which provide the right conditions. Space habitats provide these conditions, but are located in space, rather than on Earth. They are distinguished from space vehicles and stations by long-term occupancy and size. Theoretically a space habitat could be sized for a single person, but needed technical skills and psychology probably set a lower limit at about 6-8 people. At the other extreme, linked assemblies or close formations of rotating orbital habitats, or large non-rotating ones, could support planetary-scale populations. Surface habitats can range from individual pressurized modules, to permanent bases and cities. At the limit, an entire moon or planet can be converted to habitable conditions, which is known as **Terraforming**. Supporting planetary-scale populations won't be needed for a long time, so most current work is aimed at the lower end of the size scale.

Since the environment parameters that humans and agriculture prefer is fairly narrow, habitats will have similar functions regardless of size. These functions include:

  - Atmosphere Maintenance -
  - Temperature Control -
  - Artificial Gravity -
  - Lighting -
  - Radiation Level -
  - Food Supply -
  - Water Supply -
  - Waste Disposal -

> Artificial habitats are much less mass intensive than natural ones. For example, the Space Station uses roughly 100 tons to support each person, while the Earth uses about 5 trillion tons. So habitats constructed in the Solar System can either support vastly larger numbers of people, allow vastly larger living space and energy use for current population levels, or only use a small fraction of available resources.

- **Closed Life Support** - All space projects that include people require a **Life Support System** of some kind. To date, these systems have been "open" in the sense of needing outside supplies of oxygen and food. "Closed" systems recycle part or all of the materials used to sustain life, therefore the amount of stored or newly delivered supplies can be reduced. If coupled with local extraction of needed materials, outside supplies can be eliminated entirely. Water, air, and food are the principal items that would be recycled. Closed systems can be artificial, using machines and chemical processes, or ecological, using living things. For food at least, people have a preference for naturally grown items, and plants naturally produce Oxygen, so this tends to result in mostly ecological systems. Closed life support can be integrated with human living space, such as a habitat dome with both living quarters and farm areas. Alternately, a greenhouse might be optimized for plant growth conditions, including high CO_2 levels, and people use breathing equipment and remote control.

- **Habitat Construction** - Launch from Earth involves passing through the atmosphere, where large size increases drag, and very large objects are not mechanically suited to fit on launch vehicles with diameters of a few meters. The Space Station was therefore assembled from many smaller prefabricated components over time. Even larger future projects may use multiple orbital construction methods. Folded structures, such as solar panels, have been extensively used for decades. To contain an atmosphere, rigid pressure vesels have been used in the past. Inflatable structures, which are collapsed for launch, are currently being demonstrated. In the future, large habitats may be assembled from components launched in compact form, or manufactured locally in space.

- **Recycled Vehicles** - A conventional rocket takes the final stage, along with the payload, into orbit. By re-fueling the stage, or by converting the stage tanks and structures to another use (such as an occupied pressurized module), some payload weight and volume is saved. For example, the Skylab space station was made from a converted Saturn V 3rd stage. The conversion can be done before launch, as in the case of Skylab, preparations for later conversion can be installed, such as connections for refueling and attaching later hardware, or the stage can be entirely unmodified, and all the changed performed in orbit. A number of studies were done on re-using Space Shuttle external tanks for purposes like pressurized living space or raw aluminum for orbital manufacturing, but these did not progress to actual projects. Re-fueling of upper stages has also been studied, but not tested.

Transport

Space transport began before reaching orbit, in the form of ballistic missiles. It continues to be needed today, for delivery of new equipment and facilities, cargo, and raw materials. The primary transport method used to date is chemical rockets, both for launch from Earth, and in-space missions. More recently, air-breathing propulsion using carrier aircraft, ion-electric, aerobraking, and gravity assists have been used. Solar sails, electrodynamic, and higher speed air-breathing engines are at an experimental stage. Existing and future transport methods are more fully described in Part 2 of this book. These methods will continue to be needed as long as space projects exist. Some future transport applications include:

- **Hazard Removal** - Artificial and natural hazards exist in space and on Earth, and transportation methods can remove them or place them in safe orbits. Types of artificial hazards include **Space Debris** from old satellites, empty rocket stages, and fragments thereof. Below about 2000 km altitude, this debris is denser than natural small meteoroids. Transporting nuclear wastes and very hazardous biological materials from Earth to safer locations in space has been studied, but is currently too expensive. Debris impacts can damage active satellites, where the latter are expensive for the same reason, so there is a stronger economic case for removing the debris.

  Asteroids and comets are known to hit large bodies everywhere in the Solar System, including Earth. Evidence includes impact craters on every solid body, a comet impacting Jupiter, and current falls on Earth. A future project is diverting asteroids and comets which are on dangerous paths. Extensive searches have been ongoing for Earth-approaching objects, and their discovery rate is increasing as better telescopes are built for this purpose. Lunar impacts are often neglected, but more mass can be tossed off the Moon, because it's smaller, to end up in Earth's larger gravity well nearby. You be killed just as well by a 1 ton Lunar fragment as by a megaton asteroid, but the deaths are more distributed in time and space. Methods to move or destroy dangerous objects is at an early experimental stage.

  Long period comets are undetectable with current technology until they get within about 10 AU of the Sun. If one was headed towards Earth, there is not time to arrange a shift in it's orbit, so the only reasonable way to deal with it is to use an interceptor with one or more large nuclear bombs to fragment or destroy it. Comet trajectories are hard to predict because they have natural rocket thrusters in the form of gas jets. Future projects may place search telescopes farther out, to find such comets earlier. They may also station interceptors farther out into the Solar System, so fragments have more time to disperse. If more time is available, dangerous comets can be diverted using their own material as propellant, or another, smaller, natural object can be diverted to collide with it.

- **Interstellar Transporter** - The energy to transmit the description of an object to another star, even at an atom by atom level, is about a million times less than the energy to physically move the object from one star to another. In the far future, after a first probe sets up a receiving/replication station at the other star, other objects are more efficiently scanned, transmitted, and reconstructed at the receiving end. Using atomic scale technology (such as scanning tunneling microscopes) it may be possible to eventually scan and send people this way. The subjective time to travel at the speed of light is then zero, although the actual transit time is still governed by the speed of light.

Science and Exploration

Science and exploration have been the main motivations for missions beyond Earth orbit, once the **Space Race** of the 1960's wound down. They have also been important motivations in Earth orbit, but no longer the dominant ones. Most work in these areas is paid for by government agencies, so their pace is limited by available budgets. Humans have not traveled past Low Earth Orbit since the Apollo program, but numerous **Solar System Probes** have visited all the discovered major planets, some asteroids and comets, and five have left or are on trajectories to leave our home system entirely.

Future human exploration of the Moon and Mars has long been discussed, but has mostly been held up by lack of sufficient funding. Current development of the large **Space Launch System** (SLS) rocket, lower cost commercial vehicles, and better technologies, including use of in-space resources, is starting to change the situation. With limited funding, divisive arguments about "Moon vs Mars" and "Humans vs Robotic" have stalled program planning. In reality, these are false choices. With better technologies, multiple destinations are affordable. Robots can prepare the way for and assist humans. Leveraging commercial systems brings economies of scale which agency budgets can't reach on their own. The proper way to think of it is science and exploration pave the way for future commercial and industrial activity, and commerce and industry make the science and exploration more affordable. Part 4 of this book discusses both types of projects in the context of an integrated long-term program.

There are many planned astronomical instruments and interplanetary probes, and many more have been proposed, but are too far in time to be funded yet. Only a few human missions beyond Low Earth Orbit are in development or detailed planning. Detailed plans for Lunar or Mars missions have not yet been prepared, although technology work is in progress.

Communication

The use of space for communications is well developed. It is the primary purpose of 52% of the 1261 operational satellites as of the end of 2014. Essentially all other satellites have some communications functions. If costs were lower, use of space communications would increase in areas like satellite broadband. This isn't used much today because of higher cost and technical limitations compared to ground services. A future large network of low-orbit satellites could overcome some of the limitations and provide world-wide coverage. To date, most satellite communications has been by microwave radio. Some experiments have been done with laser transmission, which can provide much higher bandwidth, especially for locations beyond Earth orbit.

- **Gravity Lens Relay** - Massive objects like stars bend light via gravity. If you travel a sufficient distance from the star, at least 550 AU for the Sun, that light reaches a focus. A far future project would be to use the star as a giant lens, to focus communications over interstellar distances. For optimal signal relay, you would use such gravity lenses between pairs of stars, and boost the signal at each star before passing it on to the next. Such a network could link the entire galaxy, although the speed of light presents a significant obstacle to practical use. Gravity lenses can also be used for astronomy, which requires the least effort as you only need equipment around the Sun, and for power beaming to distant spacecraft. The latter requires much more effort than communications, since the power levels are much higher

Entertainment

This category includes activities like **Space Tourism**, **Zero-Gravity Sports**, and other activities purely for entertainment. Tourism on Earth represents about 3% of worldwide spending, and involves over a billion annual travel arrivals. There is every reason to expect a lot of people will travel to space for entertainment, if only the extreme cost and significant risk were reduced. Spectator sports in zero gravity may come sooner than general travel, because top athlete income is already in the same range as astronaut launch cost to orbit. Racing or prize competitions are a possibility, since events like the **America's Cup** already draw huge budgets and involve high technology. Such competitions would also have the useful purpose of promoting improvements in space technology, as the **Google Lunar X Prixe** is demonstrating.

Part 2: Space Transport Methods

In Part 1 we discussed the basics of the relevant sciences, including propulsive forces (Section 1.3) and energy sources (Section 1.4). These two items can be combined into one or more space transport methods. In Part 2 we will discuss the numerous transport methods, including those in current use, in various stages of development, and proposed or theoretical ones. Before looking at individual methods, we will briefly look at why reaching space has been so difficult and expensive in the past, and what can be done in general to change that situation. We also present some reference tables listing the various methods.

The Space Transport Challenge

An ideal transport system has an infinite life and minimal operating cost when delivering a desired cargo. Real systems never reach that goal, and the designer's job is to approach it as closely as possible given the level of technology and other project constraints like funding and schedule. For air and space transport, the desired cargo is called a **Payload**. For transport purposes it is measured in absolute mass (kg) or **Mass Fraction**, which is the percentage of payload mass to total vehicle mass including payload.

The Historical Problem

The difficult job of the mid-to-late 20th century rocket designer was to find the best compromise between high cost and small payload when going to Earth orbit. This compromise was forced by three factors which conspired to produce a small payload fraction, and therefore high relative cost:

- The mass and size of the Earth, which determines the velocity needed to reach orbit,
- The energy contained in chemical rocket propellants, which determines how much of them you need, and
- The strength and other properties of the materials used to build the rocket, which determines the mass of the vehicle hardware.

It happens that the combination of these three factors produced a small, or even negative payload in an ideal one piece, long life rocket. The total launch mass of such a rocket consists of 3 main parts: Propellant, Vehicle, and Payload. Payload mass is what is available after accounting for the first two, and in 20th century designs that remainder could be negative, or at best a few percent. For example, a single stage LO2/LH2 vehicle with a mission velocity of 9000 m/s and an exhaust velocity of 4500 m/s might be 86.5% fuel, 10-15% vehicle hardware depending on operating life, and therefore -1.5 to +3.5% payload. The lower hardware weight was associated with more flimsy single use construction, and the higher weight with thicker long life design. LO2/LH2 was the highest energy fuel combination. The large ratio of rocket to payload weight led to high launch cost measured in $/kg.

Various compromises had to be made from the ideal concept of a vehicle you could just refuel and launch again like an airplane. One was to make the hardware last just a single flight, thus allowing lighter structures than ones built to last many flights. Another is to drop parts of the vehicle during flight, producing a **Multistage Rocket**. As fuel is used up less thrust is required to maintain acceleration, so fewer or smaller engines are needed. Tanks also get emptied, so you can drop the excess engines and tanks once you don't need them any more. The remainder of the vehicle starts anew to accelerate towards orbit, but has the benefit of the velocity gained with the previous stage, and less vehicle weight. Low service life and staging are are expensive - you have to replace or re-assemble the hardware, but were necessary given the state of technology. The designer had to find the best balance between high cost due to small payload, or high cost due to discarding or rebuilding a complete rocket. Single-use, multi-stage rockets are still the most common method used to reach Earth orbit, and so space transport to date has been very expensive.

21st Century Solutions

The mass and size of the Earth and the energy in chemical propellants are fixed, and structural materials improve fairly slowly. Significant improvements in cost can come from using the latest materials, optimized fuel choices, and recovery and reuse of vehicle parts. But ultimately chemical rocket designs are limited by orbit velocity and fuel energy, which are fixed values. To make more dramatic changes in cost we therefore must break one or more of the 20th century's limiting factors. For example:

- You can build large structures that span part of the Earth's gravity well. This reduces the velocity the vehicle needs to provide. Even a small reduction shifts the near-zero payload fraction into positive territory. Past systems have implicitly assumed chemical rockets do the whole job of reaching orbit, because it was the only type of propulsion considered useful for that job. There is no physical law that demands using one type of propulsion for the entire job, and using multiple methods often gets better total performance.
- There are now a multitude of alternatives to traditional chemical rockets, ranging from some in actual use, to those merely theoretical as yet. Part 2 of this book attempts to list all the known ones, of which there are 83 so far, not including variations. Many of them have better performance, reducing the fuel required and increasing the net payload.
- Since the mid-20th Century better structural materials and lighter weight components have emerged, but some current rockets still don't take advantage of them.

Although having more options is more complex, the modern designer should consider the full range of available transport methods, and apply them where they function best. The potential gain in going from short life, low payload transport to long life, high payload designs more than justifies the extra work.

Beyond Low Earth Orbit

From an orbital mechanics standpoint, **Low Earth Orbit**, (LEO) is the altitude range below 2000 km, or about 1/3 of the Earth's radius. In this region the Earth's gravity is more than 58% of the surface value, and orbit periods are less than 50% higher than the 84 minutes of a theoretical orbit at sea level. From a practical standpoint, LEO is the altitude range of 200 to 1,000 km. This is above significant atmospheric drag, and below the Van Allen Radiation Belts. The historical space transport problem was to get from the ground into this orbit range. Travel beyond LEO has not been as constrained to chemical rockets as initial launch. That is because for vertical rocket launch you need to be able to accelerate the whole vehicle at more than one gravity (9.8 m/s^2) in order to take off at all, and maintain a relatively high thrust to prevent hitting the ground again while still below orbital speed. One of the key features of chemical rockets is their very high thrust to mass ratio. This made them attractive despite their low absolute efficiency.

Once in a stable orbit, lower thrust levels can be used to travel further, since you are not in danger of immediate re-entry. So alternatives to chemical rockets have been examined for those missions, and some of them even put to use in recent decades. There are now more alternatives for going beyond LEO, which give better performance or wider design choices. The larger change, however, is in the first step of getting to LEO, which was limited to one relatively low performance method until recently.

Known Transport Methods

As a starting point for designers, Figure 2.0-1 lists the 83 main space transport methods known to the book's authors. They are described in more detail in the following sections and are organized into logical categories by type. The order listed here and later in the book is by similarity of type, and not by feasibility or development status. Those factors are considered later. Many of these methods also have variations in the concept or application. From this list, the designer can then narrow the choices to the relevant options for a particular project, and eventually the final selection. By starting with all of them, you can be sure no viable option is missed.

Figure 2.0-2 places the same transport methods in a table by their numbers, since including the names makes the table too large. Rows in the table indicate different energy sources, which are grouped by major types. Columns indicate how force is applied to generate motion. They are grouped in two main categories: (1) expelling a material, which generates an opposite reaction force by Newton's Law, and (2) the vehicle interacting with outside materials or forces. The table is an aid to organizing and thinking about the methods. Some theoretical methods are not shown on this table, since we do not have definite designs for implementing them. Some of the methods could span more than one box if all their possible variations are considered, but we have assigned them to single boxes in this version. Empty boxes can stimulate thought about possible transport methods which are still unknown, or whether that particular combination truly lacks possible use.

These two figures represent the state of knowledge in mid-2012. There are likely some additional methods not yet known to the authors, and new ones will likely be devised in the future. If a serious project is contemplated, designers should survey the relevant specialists and literature to include the latest concepts.

List of Space Transport Methods (24 May 2012)

Static Structures
1. Large Towers
2. Space Elevator
3. Aerostat
4. Low-Density Tunnel
5. Magnetically Supported Structure

Dynamic Structures
6. Fountain/Mass Driver
7. Launch Loop
8. Multi-Stage Elevator

Mechanical Accelerators
9. Leveraged Catapult
10. Rotary Sling

Artillery
11. Solid Propellant Charge
12. Liquid Propellant Charge
13. Gaseous Charge
14. Rocket Fed Gun

Light Gas Gun
15. Pressure Tank Storage
16. Underwater Storage
17. Particle Bed Heated
18. Reactor Heated
19. Electrically Heated
20. Nuclear Charge Heated
21. Two Stage Combustion
22. Gravity Piston

Electric Accelerators:
23. Railgun
24. Coilgun (Mass Driver)

Air-Breathing Engines
25. Fanjet
26. Turbo-Ramjet
27. Ramjet
28. Scramjet
29. Inverted Scramjet
30. Laser-Thermal Jet

Internally Fueled (Rockets)
31. Solid Rocket
32. Hybrid Rocket
33. Liquid Rocket
34. Gaseous Thruster
35. Mechanically Augmented Thruster

Electrothermal
36 Electric Rail Rocket
37 Resistojet

Photothermal
38. Solar Thermal
39. Laser Thermal
40. Laser Detonation Wave
41. Microwave Thermal

Nuclear Thermal
42. Solid Core Nuclear
43. Liquid Core Nuclear
44. Gas Core Nuclear

Bulk Matter
45. Rotary Engine
46. Coilgun Engine
47. Railgun Engine

Ion
48. Electrostatic Ion

Plasma
49. Arc Jet
50. E-Beam Heated
51. Microwave Heated
52. Fusion Heated
53. Antimatter Heated

Particle Rockets
54. Pulsed Fission (Orion)
55. Inertial Confinement
56. Alpha Emitter
57. Fission Fragment
58. Pure Fusion
59. Neutral Particle Beam
60. Antimatter Annihilation

External Particle Interaction
61. Magsail
62. External Particle Beam
63. Interstellar Ramjet
64. Interstellar Scramjet
65. External Fuel Supply

Photon Sails
66. Solar Sail
67. Laser Lightsail
68. Microwave Sail

Photon Rockets
69. Thermal Photon
70. Stellar Photon
71. Gamma Ray

Photon Gun

External Interactions
72. Ionospheric Current
73. Gravity Assist
74. Gravity Counterweight
75. Aerodynamic Forces
76. Rheobrake

Theoretical
77. Alternat Spacetime
78. Antigravity
79. Modified Newtonian Dynamics
80. Quantum Black Hole Engine
81. Tachyons
82. Warpes Space
83. Wormholes

Figure 2.0-1 - List of space transport methods.

Table of Space Transport Methods

Energy Source Is:		Force Is Applied By:											
		Reaction from Expelled Material:								External Interactions:			
		Bulk Solid	Micro-particles	Gas Flow	Combustion Gas	Plasma	Ions	Atomic Particles	Photons	Mechanical Traction	Friction	Gas Pressure	Aerodynamic Forces
Mechanical	Compressed Gas											15, 16	
	Potential Energy									1, 2		3,22	
	Kinetic Energy	45		35						8, 9, 10	76		4
Chemical	Fuel-Atmosphere Combustion			25, 26, 27, 28, 29									75
	Fuel-Oxidizer Combustion			34	31, 32, 33							11, 12, 13, 14, 21	
	Chemical Battery												
Thermal	Thermal Storage Bed											17	
	Concentrated Light												
Electrical	Power Line			36								19	
	Electric Generator												
	Magnetic Storage												
	Photovoltaic Array	46, 47		37		49, 50	48						
	Microwave Antenna					51							
Beamed	Solar Flux			38					70				
	Laser			30, 39		40							
	Microwave			41									
	Neutral Particle												
Nuclear	Radioactive Decay							56					
	Fission Reactor			42, 43, 44				57				18	
	Fusion Reactor					52		55, 58, 59, 63, 64					
	Detonation							54				20	
Matter Conversion	Antimatter					53		60	69, 71				
	Black Hole							80					

Figure 2.0-2. Transport methods by Energy Source and Force Application.

Section 2.1 - Structural Methods

We begin our review of space transport methods with the structures category. Surviving artificial structures appear as early as 23,000 years ago, in the form of a rough stone wall at the entrance to **Theopetra Cave**. The wall was probably intended to block cold winds, since that time was near the **Last Glacial Maximum**. A "tower whose top is in the heavens" is famously mentioned in Genesis 9:3, although the baked brick, stone, and tar technology of the first Millenium BC would not have been up to the task. It is only since the late 20th Century that structural materials like **Carbon Fiber** became available, that are strong enough to perform useful space transport functions. A structure only needs to be built once, but can be used multiple times. The cost per use thus goes down the more often it is used, and the longer the structural life is. This is a very different economic situation than single-use rockets, the main space transport method used to date.

Structures in General

For space transport purposes, we want to know how useful a structural material is in and of itself, rather than as part of a vehicle with other propulsion. To find that out we can derive performance measures from the material's properties. Then we compare these measures to the transport job of reaching orbit from the surface, or changing orbits.The measures are relative to a body's gravity well or orbit velocity

Gravity Wells

Conceptually, a **Gravity Well** is related to water wells with steep sides, which you must climb if you find yourself at bottom, requiring energy to get out. For large bodies, the "depth" of a gravity well can be expressed as the surface gravity times the radius of the body, which has units of m^2/sec^2. This may be thought of as climbing one radius above the surface at constant surface gravity is the same amount of work as climbing to infinity under the actual inverse square decrease of gravity with distance. For Earth, that is 6,378,000 meters at 9.80665 m/s^s (standard surface gravity) or 62,547,000 m^2/sec^2. This is derived from the formulas for potential energy and gravitational force in the Physics section:

The formula for potential energy is $-GMm/r$ where G is the Gravitational Constant, M is the mass of the large body, m is the mass of the object of interest, and r is the radius of the large body. Since gravitational force g is GMm/r^2, the potential can be expressed as $-gr$. Since generally $F = ma$, dividing g by the object mass gives the potential per mass as $-ar$, where a is the surface acceleration of gravity and r is the radius. This is a convenient form for calculations, since both surface gravity and radius are usually known for large bodies.

This derivation of gravity well depth assumes a non-rotating, uniform, spherical object. Real bodies depart from these assumptions in varying degrees. Determining the "depth" of any point on a real body can be done by corrections to the simple formula if they are small. For fast-rotating or irregularly shaped objects a numerical integration of the gravity field may be required.

Scale Length

In physics in general, a **Scale Height** *H* is a distance over which quantity changes by a factor of the mathematical constant *e* (2.718...). It is most often applied to change in atmospheric pressure with altitude due to gravity. The variation in pressure due to gravity for gases also happens in solids. So for structural engineering purposes we can also calculate a scale height for columns or length for cables. A vertical column or cable with a constant horizontal (cross-section) area has a mass m of

$$m = DAh$$

where D is the density in kg/m^3, A is the cross sectional area in square meters, and h is the height in meters. The compressive force at the bottom (for columns), or tensile force at the top (for cables), due to its own weight, is found by the usual F = ma formula, where acceleration a in this case is the local one due to gravity. Dividing by the area A gives force per unit area, or pressure P as

$$P = DAha/A = Dha$$

The tensile or compressive strength of a material, S, also has units of force per area. Equating them and solving for h then gives the maximum height or length, H, the material can sustain before failing as

$S = Dha$ becomes $H(scale) = h = S/Da$

A taller constant width column or longer constant width cable will exceed the strength of the material and therefore collapse or break. Structural limits can be reached from other causes than gravity, such as centrifugal acceleration.We therefore call the general case **Scale Length** because it is not always due to height. As examples, common steel has a strength of 275 MPa and a density of 7800 kg/m^3, and Earth surface gravity is 9.81 m/s^2, so the scale length is 3600 meters. A very strong carbon fiber/epoxy composite column has a strength of 1300 Mpa and a density of 1650 kg/m^3 and so a scale length of 80 km. Since the Earth's radius and equivalent gravity well depth is 6378 km, the gravity well to scale length ratio for this material is 80:1.

Scale length is a theoretical value like the ultimate tensile strength at which materials fail. Real designs will always use some value below that so loads are well below the failure point. We can define the **Working Length** as the scale length divided by the design **Factor of Safety**, FS. This is a ratio of ultimate stress at failure to design stress. It is based on experience and detailed understanding of how materials fail. The chosen value is intended to reduce failure probabilities to an acceptable level for the given purpose. The **Margin of Safety**, MS, is the Factor of Safety - 1. It represents the amount of added stress above those caused by expected loads which the structure can withstand. Both safety values are intended to account for known and unknown variations in loads, and the actual strength of a structural element vs it's theoretical design strength. For example, a space structure may be degraded by impact damage or exposure to the space environment, and no longer be as stong as intended.

Efficient Design

Constant area columns or cables are simple to design and calculate for small structures, but are not an efficient design for large ones. This is because the load is a maximum at only one end of a large structure - the bottom for a column and the top for a hanging cable. A constant stress, rather than constant area, design makes the best use of the material by using all of it at it's safe stress limit. Therefore the cross sectional area must vary to suit the local loads. This results in minimum weight and cost, since you only use as much structure as you need at any point. For example, the support columns of skyscrapers are typically larger and thicker near the bottom because they are supporting the whole building and its contents, while near the top floors they are supporting much less weight above them.

For a column where the mass you are supporting is at the top, each part of the column below the top has to support that weight plus the part of the column above that point. Therefore the column has to support an increasing load as you go down, and needs a larger area to keep the stress per area the same. Similarly for a cable with a mass at the bottom, each point along the cable supports that mass, plus all the cable below that point. So the cable cross section area should increase to support the increased mass as you go up. The amount of increase in both columns and cables is 1/(working length) per meter, since the working length is over how many meters the stress will increase by 100% due to the structure itself. The fractional increase is based on the sum of supported + structural mass beyond that point. The constant fractional increase in load per meter results in an exponential taper, by a factor of *e* (2.718...) per working length. In other words the cross section **Area Ratio**, AR, where h is the total length, is

$$AR = \frac{A(bottom)}{A(top)} = e^{h/(H(scale)/FS)}$$

In theory there is no limit on the area growth, you can just keep making the structure thicker at the bottom or top. So in theory you can build a structure of any height or length with any material. In reality, the exponential growth in area from a real material implies a similar growth in structural mass and cost. At some point the design becomes infeasible to build. Where that point is depends on what the purpose of the structure is. Feasibility is especially significant for large bodies like the Earth, where the ratio of gravity well to working length is high. This produces a large exponent in the area ratio formula if you try to span the entire gravity well with a single structure.

Tip Velocity

Some large structures involve rotation about a center, rather than being vertical in a gravity well. This still results in a tapered design, but the loading force comes from centrifugal acceleration rather than gravity. The general formula for centrifugal acceleration at any radius on a rotating object is

$$a = v^2/r$$

where v is the velocity and r is the distance from the center. The distance traveled per rotation period is $2\pi r$, so the velocity of any point between the tip and center is proportional to r. Thus the acceleration at any point varies linearly with radius, and the average acceleration over the radius is half that at the tip. The same type of constant stress design as for a vertical structure leads to a tapering area from center to tip of a rotating one. Instead of a near-constant gravity along the length of a vertical structure, we have a strongly varying acceleration in a rotating one. When a rotating structure is near a large body, we have gravity forces in addition to the centrifugal ones. As the structure rotates, the direction and strength of gravity will vary relative to the centrifugal acceleration of rotation. If the structure is also in orbit, it also has centrifugal acceleration of orbital motion, which can largely cancel that from gravity.

A rotating structure made of a given material will accumulate stress from the varying accelerations from center to tip. Since the acceleration grows linearly from zero at the center, the average is half the tip acceleration. Where a(tip) x r(tip) x 0.5 equals the working length, we have the same accumulated stress as a vertical structure under constant gravity. The velocity of the tip, v(tip) is then a characteristic value for the material. We can compare the characteristic tip velocity at one scale stress to the orbit velocity for the planet or body it operates on or near. This gives us a measure of usefulness for rotating structures in that location. For example, if tip velocity = orbit velocity, you can build a device on the surface which mechanically throws payloads into orbit, or reach down from orbit at zero ground velocity if tip velocity cancels orbit velocity. As with vertical structures, you can build devices which rotate faster than their characteristic velocity, but at the cost of increasing taper factor, and thus increasing mass, relative to the supported load or payload. For a large body like the Earth, large taper factors are needed to match orbit velocity, but useful designs can be built with lower tip velocities.

Material	Ultimate Strength (Mpa)	Density (kg/m^3)	Scale Length at 1.0 gravity (meters)	Tip Velocity at 1.0 scale length (m/s)
Concrete (High Strength)	69	2800	2,513	187
Carbon Steel (AISI Type 1040, annealed)	519	7845	6,746	306
Wood (Laminated Douglas Fir/Epoxy)	55	637	8,804	349
Aluminum (AA 6061 T6)	310	2700	11,708	403
Stainless Steel (AISI Type 304, max)	1100	7930	14,145	443
Titanium (10V-2Fe-3Al)	1241	4650	27,214	614
Basalt Fiber	4800	2700	181,283	1586
Vectran Polyester Fiber	2840	1400	206,857	1694
Toray T1000G Carbon Fiber	6370	1800	360,866	2237

Figure 2.1-1. Example material properties.

Material Properties

There are a vast number of known materials, each with their own structural properties, and new ones are developed on a regular basis. **Materials Science** is the field involved in understanding, developing, and applying materials. It is a well-developed subject, which can be explored in textbooks, handbooks, and online courses, such as from MIT's **Department of Materials Science and Engineering**.

Figure 2.1-1 gives some examples of common materials used on Earth, and a few that can be used for space projects where high strength is needed. Strength is not the only important property for material selection. When doing a detailed design you should do a full search for available materials and also consider all relevant properties before a final selection. Real designs will have a factor of safety, which is not included in this table, and will also have structural overhead for items like connector fittings and coatings. The overhead can be treated as extra loads to be supported or a reduction in the useful strength of the material.

The strongest materials are fibers which are strong in tension. In order to use them for compression structures, they have to be embedded in a matrix of some other material to give them stability. Otherwise they would simply bend like a thread. An example is carbon/epoxy, which encapsulates carbon fibers with an epoxy matrix. A typical ratio by area is 60% fiber and 40% matrix. The epoxy is relatively strong as a plastic by itself, but most of the strength comes from the fibers. We do not list carbon nanotube fibers or single-crytal solids like Diamond because we do not yet have ways to produce them in large enough pieces or quantities for large space projects. They have extraordinary strength, but until they can be mass-produced, they are not yet useful for large structures.

A. Static Structures

Structural methods are divided into two main groups, static and dynamic. Dynamic ones are discussed in the next section. Static structures have parts which are mostly fixed in relation to each other. The structure as a whole may move with respect to the ground, such as the main truss of the International Space Station. Large structures are primarily governed in their design by the ratio of strength to density, or specific strength. That ratio is converted to a scale length by dividing by the local acceleration. Other important properties for a large structure include stiffness, temperature dependence of properties, and resistance to decay from the surrounding environment.

Some method to travel the height or length of the structure is often required. These methods include: **conventional elevator** (which does not need further explanation), incremental winch, **linear motor** or rails, and fluid transfer.

Incremental winch - A hanging cable, as is used with conventional elevators, that spans the entire height of a tall structure, ends up duplicating the loads of the main supporting structure and would be quite massive. An incremental winch has a small motor-driven trolley which pulls a length of cable behind it as it climbs up the structure. The cable is unreeled from a spool on the elevator compartment. The trolley then hooks the cable to a fixed point on the structure some reasonable distance up. The cargo elevator remains attached to the next lower point on the structure during this time. The elevator then reels in the spool like a winch to pull itself up from one attachment point to the next. By this method the cable length and mass are kept relatively low. The elevator car requires power for the winch. This can be by rails or wires attached to the main structure, solar arrays or other power source attached to the elevator compartment, or beamed power from an outside source.

Linear Drives - Instead of a cable, this uses either traction or magnetic force to climb the structure. Traction would use friction pressure against a rail or cable, or geared drivers against a linear toothed rail. Magnetic forces would use coils functioning as an electric motor, but instead of the coils being in a circle and producing rotation, as in an ordinary motor, they are in a line producing a linear motion. Magnetic Leviataion (MagLev) trains work this way. As with the winch, it needs an power source such as wires or conductive rails.

Fluid transfer - Rather than moving an elevator compartment, a pipe could be used for higher volume transfer. Dr. Dana Andrews, formerly with Boeing, suggested pumping Oxygen gas generated on the Lunar surface up to the Lunar L2 point on a Lunar space elevator. A column of Oxygen at .1 atmosphere at L2, and a temperature of 1000 K (a solar heated pipe can be used to keep the gas hot) would have a pressure of 2310 atm (234 MPa) at the bottom. So a single pressurized pipe section puts heavy loads on the design. A better approach is to have pumping stations spaced along the elevator, to keep the pressure rise at each station low. A pipe could also serve as a pneumatic system to transport cargo besides gases. The depth of the gravity well will determine the practicality of this method.

Regardless of the method used, lifting a mass against gravity or centrifugal acceleration takes energy according to the change in potential energy = mah, where m is mass, a is acceleration against which you are lifting, and h is the height. For example, a 2000 kg passenger elevator climbing 10 meters per second in Earth gravity requires 2000 x 10 x 9.81 = 196,200 Watts.

1. Towers

Alternate Names: Space Tower, Megastructure

Type: Potential Energy via Mechanical Traction

Description:

Towers are self-supporting compression structures, whose main purpose is not habitation. When the main purpose is to house people, we call them tall buildings, or **Skyscrapers**. The **Eiffel Tower** is probably the most famous example, but they are also commonly used as transmitting towers, such as the **w:Tokyo_Skytree** (Figure 2.1-2). As noted in the example material properties table above, advanced aerospace materials have scale lengths of many kilometers, so it is possible to build towers in that size range. Such towers can be used as a high altitude astronomical platform, a launch platform for a propulsive vehicle, or a support structure for an accelerator system.

Design - In theory a tower of unlimited height could be built. At some height, though, the exponential growth of the base area and total structural weight and cost makes it impractical to go higher. For example, let us assume a structural factor of safety of 2.5, and that a launch tower is used many times. The tower mass might then be limited to 100 times the rocket and equipment mass at the top. A carbon fiber/epoxy tower on Earth would then be limited to about 150 km in height.

In a real structure, the load, the mass you are supporting besides the structure itself, probably won't all be at the top. Design calculations then have to account for where the loads are distributed along the height. Additionally, for the bottom 20 kilometers or so on Earth, wind loads, ice build-up, and other environmental effects have to be accounted for. Above 20 km, ultraviolet light and atomic oxygen can attack certain structural materials. This is not commonly a problem at low altitude, so you need a protective layer for the structure or choose different materials.

A large tower would typically be built as a truss, like the Tokyo Skytree example. The space between the vertical elements in a truss gives it stability, but it does not have to be a solid structure to support most reasonable space-related loads. Truss elements will bend if too much load is placed along their length - imagine pressing on the ends of a drinking straw or spaghetti noodle. Stiffness or **Elastic Modulus** is a material's resistance to this bending. To make best use of the material strength, the design is often made so that buckling (failure from bending) and crushing (failure from direct load) happen at the same time. For high strength materials this results in individual elements roughly 20 times longer than their smallest cross section. It also results in the tower as a whole being roughly 20 times taller than the base width. These are only generic values, the real ratio would be determined by structural analysis of the actual design conditions.

Figure 2.1-2. Tokyo Skytree.

Construction - These types of towers can be built 'from the top down' in order to avoid human construction work in a vacuum. In this process, the top section of the tower is assembled at ground level. Hydraulic jacks then raise the tower up by one section length. The next section down is then installed underneath the top section. The progressive jacking process is repeated for the whole tower height, so all the construction work takes place near ground level. Special anchoring provisions are required to stabilize the tower while being built in this fashion. Since the tower is typically tapered, anchor masses, jacks, and assembly cranes must gradually move outward from the center as the tower grows. If remote controlled robots are used for construction, then the standard method of building from bottom to top can be used. Wind loads are significant below 20 km altitude, where the product of atmospheric pressure and wind speed produces maximum **Dynamic Pressure**. To reduce these wind loads, the structural elements can be enclosed in pivoting airfoils, which have a much lower **Drag Coefficient** than circular or triangular struts.

Status:

The tallest man-made above-ground structure is the **Burj Khalifa**, in Dubai at 830 m (2723 ft or 0.51 miles). The tallest freestanding structure is the Magnolia **Tension Leg Platform**, which is 1580 m (5200 ft) from the sea floor to the top of the surface platform. The tallest building under construction is the

Jeddah Tower, in the Saudi Arabian city of the same name. It is estimated to be 1000 meters (0.63 miles) high when completed. A number of larger **Tall Buildings and Structures** have been proposed but not yet started construction. Civil engineering and construction are very well developed fields. Suitable materials exist for multi-km tall towers, though they may require a advances in construction techniques. The economics of such towers is more limiting at the present time.

Variations:

- **1a Unguyed Tower** - This type is self-supporting from its base, like a **Pine Tree**. Like trees, such towers may need an extensive underground structure to distribute the weight and tipping forces like wind. The base diameter will typically be 5-10% of the height to prevent bending. In the lower part of the tower, wind loads may require the base to spread more than the upper part, which only depends on buckling for its necessary width. This approach assumes that most of the loads on the tower are vertical, as in an elevator riding up and down the tower height. Carbon-epoxy materials can support up to 700 MPa in compression. The largest rocket in development is about 2.55 million kg in mass. If we allow 10 million kg total for an example use like rocket launch platform on top of the tower, the load is about 100 MegaNewtons (MN). This load can then be supported by 1/7 of a square meter of cross section of carbon-epoxy columns. This is far less than the minimum launch platform size of about 2500 square meters. So the columns can be hollow tubes, and the tubes spaced apart from each other in an open truss (Figure 2.1-2). The spaces between the vertical structural elements can be used for other purposes.

- **1b Guyed Mast** - Masts are structures stabilized by diagonal supports like ropes or wires, as in the **Rigging** of sailing ships. Diagonal supports can better resist sideways forces like wind. This is commonly done with television and radio towers because the antenna itself is not very heavy, and so the main loads are winds on the lightweight tower structure. A very tall structure may combine diagonal supports in the lower 20 km, where wind is significant, and be self-supporting in the upper portion.

- **1c Series of Towers** - An electromagnetic accelerator for people and delicate cargo may be hundreds of km long, with the upper end many km high to avoid aerodynamic drag and heating, forming a long ramp. To support the device you can use a series of towers of increasing height as supports, with connecting structure similar to a suspension bridge between them.

- **1d Inflatable Tower** - Many materials are stronger in tension than compression, so concepts like the **ThothX Tower** have been proposed, using internal gas pressure to put the structure in tension and support loads.

References:

- Krinker, M., **Review of New Concepts, Ideas, and Innovations in Space Towers**, 2010.

2. Space Elevator

Alternate Names: Skyhook, Beanstalks, Jacob's Ladder, Space Bridge, Geosynchronous Towers, Orbital Tethers

Type: Potential Energy via Mechanical Traction

Description:

In 1895, astronautics pioneer **Konstantin Tsiolkovsky** was inspired by the recently built Eiffel Tower. He envisioned a tower that would reach all the way to **Geostationary Orbit**, where orbit period matches the Earth's daily rotation. We now refer to any similar concept as a **Space Elevator**, because you could use an elevator rather than a rocket to reach space. Since the top end matches local orbit velocity, you can merely let go and are in orbit. Tsiolkovsky's concept was more a thought experiment than a practical design, because in 1895 no materials existed with anywhere near the required strength

A more modern version of the concept (Figure 2.1-3) assumes a cable in tension rather than a tower. This is because fibers like PAN-derived carbon are currently the highest strength available bulk materials, and bulk quantities are needed to build a space elevator. The part of the elevator below stationary orbit sees more gravity than centrifugal acceleration, and would fall down if unsupported. Therefore this version requires a counterweight above that orbit, where centrifugal acceleration is higher than gravity, and the weight then pulls upward on the attached cable. This version of the space elevator is often used in popular illustrations, but early 21st century materials are still not strong enough to make this size elevator practical. Smaller versions with shorter cables, however, can result in feasible designs. We can call these **Staged Elevators** since they only provide part of the velocity to reach orbit, in the same sense that the stages of a multi-stage rocket do. A different transport method or multiple elevator stages are needed to do the full job of reaching orbit.

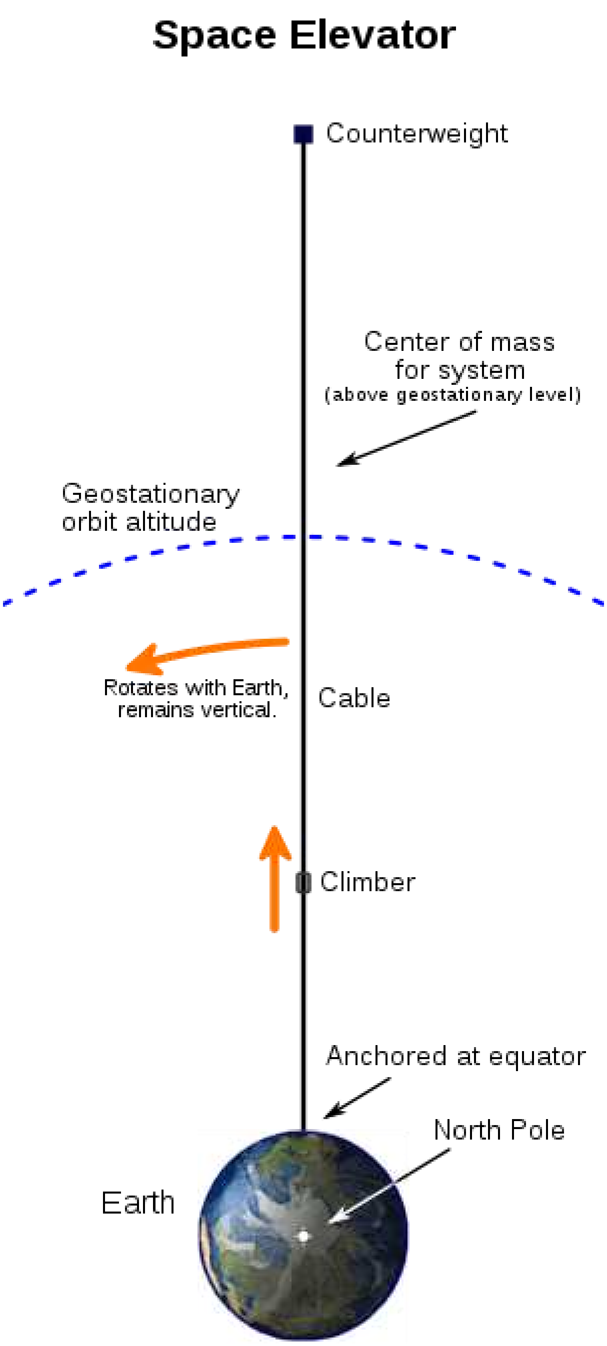


Figure 2.1-3. Basic space elevator concept.

If staged elevators are not connected to the ground, they are not required to be centered on a 24 hour orbit, nor rotate at the same rate the Earth does, creating a whole class of possible designs. They can be located in whatever orbit is needed, and will appear to be in motion when viewed from the ground. For bodies smaller than the Earth a full elevator is more feasible, because the lesser gravity well can be spanned with existing bulk materials. If future materials about 3 times stronger than current carbon fiber become available, the original single-stage elevator would become feasible from a structural materials standpoint. It would still be a very large construction project, and have other technical challenges to overcome. The rotation state of the elevator can be vertical, which is one rotation per orbit when seen from inertial space; swinging, where it varies by some angle from vertical but does not do a full rotation relative to the ground; or rotating where it does rotate relative to the ground. The rotation sense can be forward, where it is the same direction as the orbit around the planet, or backward where it is opposite. Normally it would be forward, since that results in lower velocity at the bottom end relative to the ground, and higher velocity at the upper end for injection into transfer or escape orbits.

For any type of space elevator, the structure can be used multiple times over it's design life. So the construction cost is divided by the number of times it is used. The larger the structure is, to gain more performance, the higher the cost. When maintenance and operations are added to the construction cost, there will be an optimum size and cost for a given traffic level. Mass and cost grow exponentially with size and performance, and construction cost per use only decreases as the inverse of the number of uses. So at some point the lower cost from many uses is overcome by the increasing mass and cost of a larger structure no matter how many times it is used. Therefore economics is a limiting factor on large space elevators.

Gravity and Stress - Where a tower has to resist the compression force created by gravity against the solid surface of a body, a space elevator needs to resist the tension force created by the difference in gravity between it's parts, or from it's own rotation. Structural elements can store and transfer momentum and potential energy to vehicles or cargo, and support objects away from the structure's center of mass or under rotation. Structures in orbit will naturally tend to align vertically (Figure 2.1-4, right) because gravity forces decrease as the square of distance. So the lower end of a vertical cable has a lower potential relative to the center by more than the higher end sees a higher potential. The total energy of a vertical cable is then less than that of a horizontal one, so it tends to fall into that state. The difference in gravity (otherwise known as tides) provides tension to keep the structure extended.

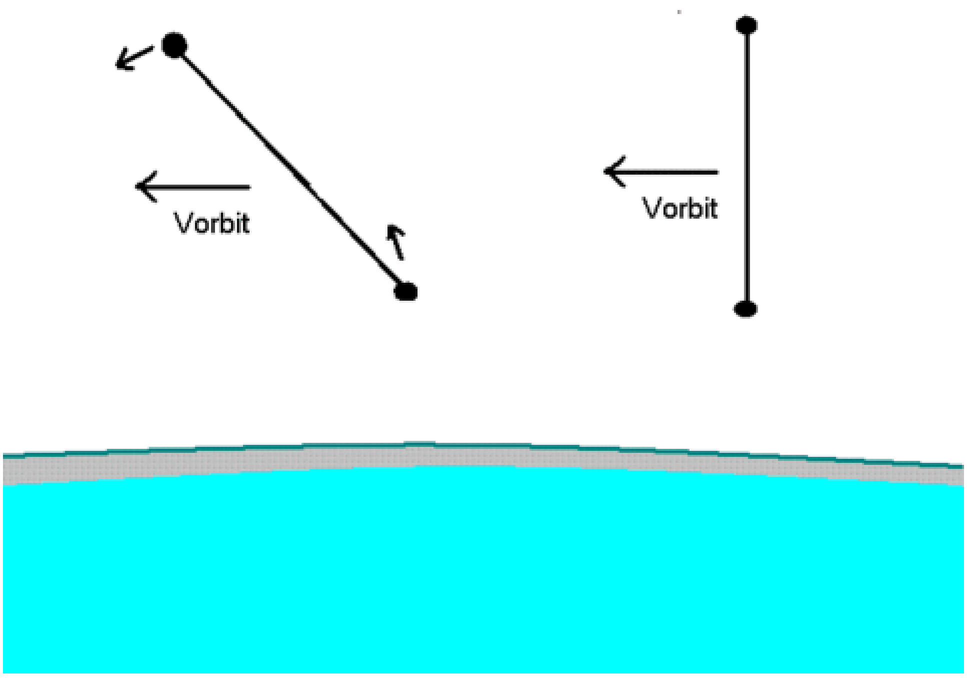


Figure 2.1-4. Staged elevators: rotating (left) and vertical (right).

Staged elevators can be built any size up to the practical limit of the materials used. In the lower limit of zero length a space elevator reduces to a simple object in orbit. For intermediate lengths and vertical orientation, the velocity of the bottom end is the velocity of the center times bottom/center distances r from the center of the body. This simply reflects that the center and bottom travel paths of length 2πr in the same orbit time. The bottom's velocity can be further reduced if you rotate the entire cable such that the bottom end travels opposite the orbit direction relative to the center (Figure 2.1-4, left). The tip's maximum rotation velocity relative to the center is then governed by the mass and cost of the structure growing exponentially as you make it more capable. For the Earth, and current bulk materials, it is not practical to counter all of orbit velocity with rotation, but about half can be. On smaller bodies the entire velocity can be countered by rotation, so that the bottom end momentarily comes to rest relative to the body's surface.

The remaining velocity of the bottom end relative to the surface has to be provided by some other method. When a body rotates, like the Earth does, some of the remainder can be supplied that way. In our planet's case, equatorial rotation velocity is 465 m/s, or 5.88% of orbit velocity. The roughly half of orbit velocity not supplied by the elevator's or the Earth's rotation is a substantial improvement over needing to supply all of it. The difficulty of building a transport system like a chemical rocket is also non-linear with velocity. For example, if it could provide 3.5% of launch mass as payload to orbit doing the whole job, it could increase payload to 27% if it only needs to do half, an increase 7.7 times. By using a rocket + elevator as a two-stage system, the sum of smaller mass exponents is less than each exponent by itself, so the total cost and difficulty is less using both than using either one for the entire job.

The rocket stage does not place itself fully in orbit. It docks temporarily at the tip of the cable, and only the cargo travels further using the elevator. This is less total energy per delivery than if the rocket stage hardware had to travel all the way to orbit. The elevator cable serves as a **momentum bank** to store orbital kinetic energy, which can be transferred to the payload. The kinetic energy can be stored by running an electric propulsion system, which is much more efficient than conventional rockets. Using an elevator stage lowers the overall difficulty of reaching orbit, and can lower the overall cost if well designed. Material strength to density ratio is the critical criterion for designing these types of transport systems. Their mass is highly non-linear with strength because doubling the strength reduces the exponent part of their mass ratio by half.

Structural Dynamics:- The forces affecting a space elevator design vary with time, so they are dynamic forces rather than static ones. This includes arriving and departing vehicles, internal movement of cargo along the structure, deployment of extension cables if used, thrust for orbit maintenance and rotation, varying gravity in an elliptical orbit or if the elevator is rotating, varying tidal forces from the Sun, Moon, or planetary satellites, and thermal stress from going into and out of shadow as it orbits. Elevator designs can be truss-like, with sufficient compressive structural elements to keep a stable shape against these varying forces. They can also be cable-like, with primarily tension elements, and the structure allowed to flex with the applied forces, or they can be a mixture of the two. Active damping for vibrations can be applied with shock absorber/spring combinations or with thrusters along the structure. Design for dynamic forces is similar to the design problem of a suspension bridge, which must withstand static forces from its own weight, and dynamic ones from vehicle traffic and varying winds. In both cases, the design must not exceed safe working stresses at any point in the structure, under any combination of forces.

Because of the typically high slenderness ratio (ratio of length to maximum width) and varying forces noted above, the structural dynamics will be complex and require a good theoretical understanding and likely computer simulation. A further complexity is unlike terrestrial skyscrapers, which are constructed empty and then loaded when complete and not usually changed afterwards, a space elevator may grow over time while already operating. Thisis likely because a large elevator can assist in it's own construction by reducing the work for a launch system, and can help offset it's cost by operating as soon as possible. So instead of analyzing a completed building and then checking construction loads do not exceed design loads, a growing elevator would need analysis over a continuous range of sizes.

Maintenance and Repair - Space elevators, like large structures on Earth, are subject to environmental degradation and occasional sudden damage. These include atomic oxygen, electrostatic discharge, solar UV and high energy particles, trapped radiation belts, impacts from natural meteoroids and human-made orbital debris, and accidental vehicle collision or pressurized system failure. For long-term operating life and safe operation, a space elevator has to be designed for all these causes, with a program for maintenance and repair. In case of catastrophic damage, the design should minimize risk to people and property.

Status:

The **Elevator** on Earth dates as early as Archimedes in 236 B.C. The modern safety elevator was introduced in 1852 by Elisha Otis, and a descendant of his company is still the largest manufacturer of vertical transport. Ropes and cables have long been used on Earth for hoisting loads and stabilizing tall structures such as ship masts and transmitting towers. A space elevator refers to a complete transportation unit. A **Space Tether** refers to a cable in space that has a number of possible uses, one of which is the structural element of a space elevator. A number of **Space Tether Missions** have been flown as experiments, but not yet as an operational transportation system. The largest rigid structure in space to date is the ISS **Integrated Truss Structure**, which is 108.5 meters long. Taller structures on Earth await economic reasons to build them, they have not reached the limits of available materials. Space elevators require enough traffic to similar orbits to justify their construction. Significant traffic exists as of 2016 to synchronous orbit, but not enough to justify a synchronous-capable elevator system.

A **Variable Gravity Research Facility** has been proposed in Earth orbit to study the effects of partial gravity on people and growing plants. Such partial gravity levels are found on the Moon and Mars, and it is currently unknown what gravity level is necessary for health and plant growth on long missions. Such a facility would include large space structures to generate artificial gravity, and can also serve as a test bed for tether dynamics and operations. It would be a step towards required knowledge to build an operational space elevator system.

Variations:

- **2a Orbital Vertical Elevator** - This is the most commonly presented space elevator concept. It has a cable kept vertical in smaller versions by tidal forces, or in larger versions by sufficient cable or counterweight past GEO to apply tension to the part below GEO. Mass grows exponentially with gravity well depth. Therefore a compressive tower built up from the ground meeting a cable from above results in lower total mass, because it splits the structural task into two smaller exponentials. Despite that, current materials are not sufficient for a full vertical elevator on Earth. They are for smaller bodies such as the Moon or Mars.

- **2b Momentum Transfer Slingshot** - If a payload is released from the end of a vertical elevator, the other end of it's orbit will be changed about 7 times the initial distance from the elevator center of mass. This is because while attached the payload is forced to move at a different velocity than a free object would at that altitude. Once released, it then follows the free orbit defined by its release velocity. This variation increases the orbit change by adding partial rotation and dynamic extension of a cable.

  A variation from vertical cables is the orbital slingshot. This would take advantage of the tendency of a long object to auto-rotate from horizontal to vertical orientation about the center of mass due to "tidal" effects. A relatively light-weight vehicle, launched conventionally, would dock with a much more massive "orbital momentum bank" (largely consisting of discarded rocket stages left at the bank with each launch), and be hooked to a cable reel. The vehicle would be pushed out to a somewhat higher orbit, where it would fall behind the momentum bank, with cable being paid out at a matching rate. After sufficient tether has been paid out, it would be braked to a halt, putting it under tension. The momentum bank would slow and fall inward, while the vehicle would be accelerated and fall outward. It is released at the desired orientation and velocity to transfer to a higher orbit. Unlike an "elevator" system, the tether need not be long enough to continuously reach the ultimate orbit, as the vehicle will be "slung" outward up to 14 times the cable length.

  The momentum bank loses velocity in this maneuver, but could use highly efficient solar-powered electric engines (plasma, ion, or magnetic) to recover that loss over an extended period. Multiple momentum banks can be used in series the achieve higher orbits or greater final velocities. If the momentum bank uses an elliptical orbit (cheaper for a rocket launched vehicle to intercept), it may be possible to insert objects into near-circular orbits by slinging at apogee. The vehicle can also take on fuel at the momentum bank, as the empty rocket stages already have propellant tanks that can be re-used. The slingshot approach has a moderate size and velocity capacity.

- **2c Orbital Rotating Skyhook** - This is an idea devised around 1980, where a cable is kept in tension by sufficient rotation rate. On smaller bodies, where the cable end can dip low enough to grab cargo and lift it to orbit, the Skyhook name is generally used. For Earth reaching that low is difficult because of the high tip velocity and mass needed. Instead a vehicle coming from the ground provides enough velocity to meet a slower rotating tip. This version is often called a **Momentum Exchange Tether** or **Rotovator** (rotating elevator). The momentum being exchanged is between a vehicle/cargo and the and the elevator system. Again, both launch vehicle and Skyhook mass ratios are exponential in velocity, so splitting the job lowers the overall difficulty.

- **2d Atmospheric Elevator** - For this concept an aircraft or balloon/airship uses a cable to lift an object to altitude, after which it continues to orbit by other methods. With an aircraft this can be a simple tow cable where one vehicle pulls another, or a cable system which dynamically snatches and accelerates a vehicle, possibly tossing it higher than the aircraft flies. It requires less modification of the towing aircraft and not having to deal with combined aerodynamics. For an airship type lifter, it avoids having to build a tower that height, although cargo mass is relatively limited.

References:

- Pearson, J. **Konstantin Tsiolkovski and the Origin of the Space Elevator**, IAF-97-IAA.2.1.09, 48th International Astronautical Congress, Turin, Italy, 1997.
- Cosmo, M. and Lorenzini, E. **Tethers in Space Handbook, 3rd ed.**, Smithsonian Astrophysical Observatory, Dec. 1997.
- Carroll, J. **Guidebook for Analysis of Tether Applications**, for Martin Marietta Corporation, Mar. 1985.
- Multiple Authors, **Space Elevator**, search of NASA Technical Reports Server, approx 4775 items from 1916 to 2016.
- Alpatov, A. et. al., **Dynamics of Tethered Space Systems**, CRC Press, Apr. 2010.

  - **Rotating Elevator** -

- Carley and Moravec, **The Rocket/Skyhook Combination**, L5 News, March 1983
- Ebisch, K. E., **Skyhook: Another Space Construction Project**, American Journal of Physics, v 50 no 5 pp 467-69, 1982.

- Baracat, William A., Applications of Tethers in Space: Workshop Proceedings Vols 1 and 2. (Proceedings of a workshop held in Venice, Italy, Octover 15-17, 1985) NASA Conference Publication 2422, 1986.
- Anderson, J. L. "Tether Technology - Conference Summary", American Institute of Astronautics and Aeronautics paper 88-0533, 1988.

3. Aerostat

Alternate Names: High altitude balloon, Airship, Inflatable Tower

Type: Potential Energy by Aerodynamic forces

Description: This method uses lift generated by pressure and density differences but not primarily from velocity such as wing lift. One approach to minimizing drag and gravity losses for a launch vehicle is to carry it aloft with a high altitude balloon or airship. Research balloons have carried ton-class payloads in the range of 15-30 km high, which is above the bulk of the atmosphere. Another approach that has been proposed is to use pressure-supported structures of great height. The highest strength materials are strong in tension, so an inflated structure in theory can support itself. Wind loads on a large pressurized structure are a major design issue. If a less dense gas is used than the surrounding atmosphere, the structure will be partially buoyant and not require the same scaling as one that depends on compressive strength. Sufficiently large structures, which would have low surface to volume ratios, could float just from heating the interior air.

Status: Balloons, airships, and pressure supported structures have been in use for a number of years, and some experiments have been done to launch a rocket from a balloon. They have not reached orbit yet.

Variations:

- **3a Balloon Carrier** - A device producing lift and carrying an instrument package or launch vehicle, but not a propulsion system of it's own. They have been used extensively on Earth for science, and been proposed for other planets.
- **3b Airship** - A device combining buoyant lift and and some combination of aerodynamic lift and forward propulsion.
- **3c Orbital Airship** - This concept has been proposed by **JP Aerospace**. It involves a very lightweight airship which starts from a floating platform and accelerates via solar-electric thrusters to orbital velocity. It is not known if this is technically possible.
- **3d Geodesic Sphere** - A triangulated frame supporting a pressure skin can float with merely a temperature difference between inside and outside if sufficiently large. Since structure mass goes with area, and lifting force goes with volume, if built sufficiently large it will float.
- **3e Pressure Supported Tower** - Uses lift force from higher interior pressure to raise a structure. This can be generalized to pressurizing any structural element to help support it.

References:

4. Low-Density Tunnel

Alternate Names:

Type: Kinetic Energy by Aerodynamic Forces

Description: Traveling to or from a large body with an atmosphere, such as the Earth, can produce large losses from drag and heating. Aerodynamic drag has a gas density factor in its formula. This concept reduces or avoids those losses by using a lower-density gas or vacuum. Lower density can be obtained by using a gas with a lower atomic weight, such as hydrogen, or by pumping out some or all of the gas. This is not a transport method in and of itself, but rather a way to avoid losses.

Status: Low pressure pipes are a common device. It has not been tried for space transport.

Variations:

- **4a Light Gas Tunnel** - One or more light gas balloons or pipes are strung along the path of a vehicle or projectile. The gas has a lower density than air. The formula for drag is 0.5*C(d)*Rho*A*v^2, where Rho is the density. Thus the lower density will lower drag. High speed travel through any gas will develop shock waves, so the size of the projectile relative to the size of the tunnel needs to be small enough that the shock waves will not damage the structure.

- **4b Evacuated Tunnel** - An evacuated tunnel is supported up through the atmosphere by a combination of towers or it's own lift from displacing air. A launch system such as an electromagnetic accelerator fires a projectile up through the tunnel. Drag losses are minimized within the tunnel, and are low in the remaining part of the atmosphere beyond the tunnel. The top end requires some way to keep air from flowing in and filling the tunnel - such as a hatch that remains closed until the accelerator is about to fire.

References:

5. Magnetically Supported Structure

Alternate Names: Startram

Type: Magnetic Storage by Magnetic Field

Description: A static or time varying magnetic field produces a force to support a structure. For example, a series of large superconducting coils stacked so they repel each other and support a cargo. Alternately current carrying wires generate repulsion between the ground and the structure.

Status: Startram is a concept proposed using magnetic levitation, but has not reached experimental versions yet.

Variations:

References:

B. Dynamic Structures

Static structures rely on constant forces such as from the strength of materials to hold themselves up. Dynamic structures rely on the forces generated by rapidly moving parts to hold up the structure. The advantage of this approach is it can support structures beyond the limits of material strengths. The disadvantage is that if the machinery that controls the moving parts fails, the structure falls apart.

6. Fountain/Mass Driver

Alternate Names:

Type:

Description: An electromagnetic accelerator provides a stream of masses moving up vertically. A series of coils decelerates the masses as they go up, then accelerates them back down again, at a few times local gravity. When they reach bottom, the accelerator slows them down and throws them back up again, at a high multiple of local gravity. Thus the accelerator is many times shorter than the fountain height. The reaction of the coils to the acceleration of the fountain of masses provides a lifting force that can support a structure. The lifting force is distributed along where the coils are located. This can be along the length of a tower, or concentrated at the top, with the stream of masses in free-flight most of the way

Status:

Variations:

References:

7. Super-Orbital Mass Stream

Alternate Names: Launch Loop

Type:

Description: A strip or sections of a strip are maintained at super-orbital velocities. They are constrained by magnetic forces to support a structure, while being prevented from leaving orbit. A vehicle rides the strip, using magnetic braking against the strip's motion to accelerate. Several concepts using super-orbital velocity structures have been proposed. One is known as the 'launch loop'. In this concept a segmented metal ribbon is accelerated to more than orbital velocity at low Earth orbit. The ribbon is restrained from rising to higher apogees by a series of cables suspended from magnetically levitated hardware supported by the ribbons. The ribbon is guided to ground level in an evacuated tube, and turned 180 degrees using magnets on the ground. A vehicle going to orbit rides an elevator to a station where the cable moves horizontally at altitude. The vehicle accelerates using magnetic drag against the ribbon, then releases when it achieves orbital velocity

Status:

Variations:

- **7a. Birch Ring** - Consists of two parts. One is a constellation of one or more low-altitude "geostationary" satellites with a space elevator attached, at a distance from the ground that would be LEO in a normal orbit. This constellation of satellites (which can also called geostatites, since they are stationary with respect to the ground) are kept from falling despite to their low velocity by a pelletized or solid mass stream that weighs more and is moving at a rate somewhat higher than orbital velocity. The mass stream maintains a lower orbital altitude by deflection from the statite constellation onto shorter arcs. In the case of a low-mass stream, the hyperorbital path is close to a straight line, and a higher number of statites is needed to divert it around the earth without cutting through the atmosphere (in a polygon-like pattern).In the case of a high-mass stream it moves only slightly higher than orbital velocity, but has a lot more mass than the statites.

Alternate Names: Orbital Ring System (ORS), Jacob's Ladder, Skyhook

8. Multi-Stage Space Elevator

Alternate Names:

Type:

Description: A multi-stage space elevator has more than one structural element, with the parts in relative motion. For example, a vertically hanging cable in Earth orbit can have a rotating cable at it's lower end. The advantage of such an arrangement is to lower the mass ratio of cable to payload compared to a single cable. The mass ratio of a rotating cable is approximately proportional to exp(tip velocity squared). If two cables each supply half the tip velocity, then the ratio becomes exp(2(tip velocity/2)squared), which is a smaller total mass ratio. Another feature of a multi-stage elevator is that the tip velocity vector of the two stages add. Since one rotates with respect to the other, the sum of the vectors changes over time. Given suitable choices of tip velocities and angular rates, one can receive and send payloads with arbitrary speed and direction up to the sum of the two vectors. The dynamics of a multi-stage elevator are complex.

Status:

Variations:

- **8a Hanging/Rotating Elevator**- This consists of a vertical/nonrotating space elevator structure with a rotating second stage at one or both ends.This is more suited for within a gravity well, where the gravitational gradient will stabilize the first stage.
- **8b Rotating/Rotating Elevator**- This consists of two stages, both of which are rotating, to get reduced mass ratio for a given velocity. This is more suited for free space application where the lack of varying gravity across the structure will simplify the dynamics.

References:

References

Section 2.2 - Guns and Accelerators

This section covers Guns and Accelerators. Although the details are quite varied, what they share is a larger fixed installation which provides acceleration to a smaller projectile or cargo. The construction of a large fixed device is justified if you use it many times and the maintenance per use is moderate.

Note: This section is continued at Guns and Accelerators 2 due to page size.

A. Mechanical Accelerators

As the name indicates, this group uses direct force from the device structure.

9 Leveraged Catapult

Alternate Names:

Type:

Description: A leveraged catapult uses a relatively large or heavy driver to accelerate a smaller payload at several gravities by mechanical means. Such devices date back to the Medieval period, but this is an updated version using modern materials. Devices such as a multiple sheave pulley or a gear train convert a large force moving slowly to a small force moving fast, and transmit the force along a cable. The mechanical advantage produces more than one gravity of acceleration. This concept may be one of the simplest to implement on a small scale. Despite the seeming simplicity of the concept, velocities of several km/s are possible, which would greatly reduce the size of a rocket needed to provide the balance of the velocity to orbit. The performance of this concept reaches a limit due to the weight, drag, and heating of the cable attached to the payload and the magnitude of the driving force, which is divided by the leverage ratio to yield the force on the payload.

Status:

Variations:

- **9a Drop Weight** - A falling mass is connected to a vehicle by a high strength cable running over a multiple-sheave pulley, cable reels with different diameters, or connected with a gearing ratio. Two types of natural locations with large height difference are possibilities - rivergorges and mountain peaks. Locations such as the Grand Canyon and the Columbia River gorge have lots of vertical relief for the drop weight. At these locations the weight can consist of a large fabric bag filled with water from the river at the bottom. The bag can be emptied before hitting bottom. This reduces the weight that has to be stopped by a braking system. For mountain peak locations, the drop weight runs down a set of rails and is stopped by running into a body of water or running up an opposing hillside plus possibly wheel braking. The mountain location may be preferred because of the greater launch altitude. This example assumes a mountain with a solid weight sliding on rails:

We assume that a 15,000 kg cryogenic rocket using RL-10 engines is being thrown. An acceleration of 60 m/s^2 is tolerable by humans for the 20 seconds required to reach 1200 m/s assuming the human is in good health and properly supported. The linear path traversed would be 12,000 m (7.5 miles) at constant acceleration. The tow cable pulls with 900 kN (202,000 lb) of force on the rocket. High strength carbon fiber is available with tensile strengths of up to 6.9 GPa (1,000 ksi). Allowing for a factor of safety and braided overwrap surrounding the fiber bundle (to protect the carbon fibers from abrasion), we assume a working stress of 3.45 GPa (500 ksi). The cross sectional area required is then 0.00026 m^2, or a circular cable 0.0182 m (0.72 inch) in diameter

With a density of 1840 kg/m^3, the cable mass is 0.48 kg/m. For a 12,000 m long cable, the mass would be 5,760 kg without allowing for taper. Since the cable has to be accelerated also, the leading end, closer to the drive mechanism, has to be able to apply sufficient force to accelerate the rocket and all the cable in between the rocket and leading end. To a first approximation, the leading end must be 40% larger in cross section than the trailing end to account for the acceleration of the cable itself. The mean weight of the cable is then 0.576 kg/m, giving a cable mass of 7,000 kg. Something has to be done with the cable after it finishes the job of accelerating the vehicle. It can be taken care of by looping the cable back from the drive mechanism to the starting point (where the vehicle is at the start). At the completion of a launch, the loop of cable is moving at 1200 m/s, and is gradually allowed to come to a stop. Some method will be required to reduce hoop stress on the cable as it makes turns around a pulley at high velocity

The return portion of the cable has to be able to accelerate the part of the cable past the drive mechanism at 6 g's. Given a 7000 kg lead section, we get a mass of 3200 kg. Thus the total mass accelerated at 6 g's is 25,200 kg, and the total accelerating force is 1.512 MN. Given a 45 degree slope on a mountain, we assume that the drop weight accelerates down at 3 m/s^2. Thus the gear ratio is 20:1 and the force of the drop weight on the drive system must be 30.24 MN. Since a free-falling weight on this slope would accelerate at 7 m/s, there is a 4 N/kg retarding force due to the drive system. Therefore the drop weight must be 7,560 tons. If it is a 4 meter square block of steel (about the cross section of a railroad car), it will weigh 124.8 tons per meter, and hence must be 60.6 meters long. The rails must be 12,000 m/20 long, or 600 meters long.

- **9b Locomotive Driver**- A set of railroad locomotives provides the motive force, which is multiplied by a gear mechanism to a higher speed. Example: launching a 20,000 lb

vehicle at 3 g's to 1100 m/s. This requires a 20 km straight run of track. The rail cars needed would include:

1 tank car for vehicle fuel
1 special purpose car to carry the glider
2-3 cars with tow rope guides to keep the tow rope off the ground.
1 pulley system car
30 locomotives in tandem.

We assume the locomotive top speed is about 27 m/s, therefore a 40:1 gear ratio will provide the desired speed at the vehicle. Locomotive traction averages 80,000 lb/engine, or 2000 lb per engine when reduced through the gear ratio. The gear-down mechanism and launch cable drum are mounted on a flatbed rail car. This car can be anchored to a foundation on either side of the railroad track to hold it in place when the combined pull of the locomotives is exerted. The starting traction of 30 locomotives is 180 tons. Since the couplings between engines are probably not designed for this load, a set of steel cables on both sides of the locomotives are used to transmit the traction force from each engine to the gear mechanism. The vehicle is attached to the anchored rail car by a high-strength cable which is 20 km long. At 3 g's it takes this distance to accelerate to the desired speed.

Two or three rail cars are spaced out along the 20 km with towers with a pulley wheel on top, to guide the cable and keep it off the ground during the initial acceleration. The vehicle has glider type wings attached that will generate lift as it gains speed, so the vehicle will climb once it reaches 100 m/s or so. When the vehicle reaches the desired speed, the cable is released and the vehicle continues to climb under the glider's lift. Eventually the glider drops the vehicle, which proceeds under rocket power. Although 30 locomotives is a lot, you only have to lease them long enough to do a launch, after which they can return to normal railroad use.

A small prototype would consist of a single Locomotive driver. 1250 lb rocket @ 4 g peak. Final velocity = 700 m/s. Accel time = 17.5 sec distance = 0.5at2 = 6.1 km. Engine traction = 80,000 lb average @ 25:1 gear ratio.

- **9c Jet Driver** - This is similar to the locomotive case, but the gear ratio is lower since the jet can reach a higher speed on a take-off run. Example: an F-15 can tow 40,000 lb rope tension if near empty@10:1 gear ratio can accelerate 1000 lb object @ 4 g's. Aircraft top speed on deck = 300 m/s. Object top speed in theory would be 3000 m/s. In practice would be limited by aerodynamics and cable heating (perhaps to 1500 m/s? limit is not well understood)

References:

10 Rotary Sling

Alternate names: Centrifuge Catapult

Type:

Description: In principle, this is a sling or bolo scaled up and using aerospace materials. Rotating cables can reach a significant fraction of orbit velocity on Earth, or all of orbit velocity on smaller bodies. A drive arm is driven in rotation by methods such as an electric motor, or a jet mounted on its tip. In larger versions a cable with the payload attached to the end is played out gradually as the system comes up to speed. The drive arm leads the cable slightly so the cable and payload see a torque that continues to accelerate them. When the desired payload velocity is reached, the payload releases and flies off. The cable is then retracted and the drive arm slows down. When it stops, another payload is attached. In smaller versions the cargo can be connected directly to the drive arm without a cable.

In a vacuum, such as on the Lunar surface, this is theoretically a very efficient system, as the sling can be driven by an electric motor and the mechanical losses can be held to a low value. Some method of recovering the energy of the arm and cable (such as by transferring it to a second system by using the motor as a generator), can lead to efficiencies over 60% in theory

On bodies with an atmosphere, such as the Earth, the system is hindered by air drag. One method of reducing drag is to attach an aerodynamic shape to the cable material, so as to lower drag compared to a circular cable. Another is to mount the drive arm on the top of a large tower, so the cable is not moving in dense gas. A third is to generate lift along the cable or at the payload raising its altitude. The rapidly moving part of the cable near the payload then has less drag.

A counterweight is desirable in situations where the unbalanced load of the cargo would put too much stress on the pivot and supporting structure. The counterweight gets released at the same time as the cargo. Due to gyroscopic effects, the rotating cable will attempt to keep its axis of rotation fixed. When attached to a rotating body such as the Earth, the location and mounting of the cable will have to take that into account.

Design Example:

This example is for a terrestrial sling with a tip velocity of 1800 m/s, or about 24% of Earth orbital velocity, and a cargo of 10,000 kg including propulsion for going to orbit:

- Assume tip acceleration is 10 gravities (100 m/s^2) for non-living cargo.
- Then r = 32,400 meters.
- For a 10,000 kg projectile at end, cable tension is 1 MN. If carbon fiber is used with a 3400 MPa design stress, then cable area is 1/3400 m2, or about 3 cm^2.
- Cable mass is 0.6 kg/meter, adds 60 N/meter at tip. The implied scale length is 16.67 km. Acceleration falls linearly with radius, so effective radius is 16.2 km for load purposes, and cable mass ratio is 2.8:1. Therefore cable mass is 18,000 kg.
- To spin up in 1 hour, we need an average accelerating force of 0.5 m/s^2, or 5000 N.
- Drag counteracts the accelerating force. Assume the cable is shaped to be 1 cm tall by 3 cm deep. Since drag is proportional to velocity squared, and velocity goes linear with radius, drag over the entire length will be 1/3 that at the tip. At the tip drag is 0.5 Cd rho A v^2. Cd = 0.04 for shaped airfoil. rho = 1.225 kg/m3 at sea level. A = 0.01 m2. v = 1800 m/s. Then tip drag = 2381 N/m, or 1/3 x 32,400 m x 2381 N/m = 25.72 MN at sea level.
- If we assume our drive motor can produce a peak of 4 times the average accelerating force, then peak drag can be at most 15,000 N. The motor is then driven to maintain a surplus of 5000 N force above drag. Since sea level drag is 1715 too high, we want to go to an altitude where density is that factor lower, or 7.14 x 10^{-4} kg/m^3. This density is at 53 km altitude. The cable will drape in a curve due to the combination of radial and gravity forces, so the tower height will need to be approx 60 km.

Status: Slings and bolos are ancient devices. So far as is known, a modern one for space transport has not been built

Variations:

- **10a Jet Driven Sling** - A jet engine is mounted to the rotating arm in order to get high starting torque. The launch velocity is increased by the ratio of cargo radius to jet radius from the pivot point.
- **10b Two Stage Sling** - A second rotating cable or arm is mounted on the first one to reduce mass ratio of the entire system. This adds complexity, so is only an advantage where the mass ratio of a single stage version becomes too extreme.

References:

B. Artillery

Guns have been in use in the West since about 1300 AD. A "space gun" is merely one large enough, and with a high enough muzzle velocity, to be useful for space transport purposes. It also needs to point sufficiently above the horizon for the projectile not to hit the terrain, and lower the amount of air to travel through. The earliest illustration of using a cannon to reach orbit is from Isaac Newton's *Principia Mathematica*, where he was illustrating the concepts of gravity and orbits. Thus the idea is as old as orbit mechanics. In the 20th century rockets were developed for ballistic missiles and then placing payloads into orbit. They were able to reach higher velocities with lower maximum acceleration and infrastructure at the time. Even though guns have a much longer heritage, and both artillery and rockets are both still used by military forces, they were bypassed in the early part of the Space Age.

11 Solid Propellant Charge

Alternate Names:

Type:

Description: This method uses an explosive charge which vaporizes behind a projectile in a barrel. The high gas pressure generated this way accelerates the projectile to high velocity. Conventional artillery reaches speeds of around 1000 m/s. Velocities much higher than this are difficult to reach because the barrel mass becomes extreme at higher pressures, and the temperature and molecular weight of the resulting gas limits the internal speed of sound, which in turn limits muzzle velocity

Status: Artillery has a long history and extensive use. The High Altitude Research Probe project attached two naval gun barrels in series and used relatively light shells to reach higher muzzle velocities than conventional artillery

Variations:

References:

- Newton, Isaac, **Principia Mathematica**
- Verne, Jules, **From the Earth to the Moon**

12 Liquid Propellant Charge

Alternate Names:

Type:

Description: This is similar to conventional solid propellant artillery except liquid propellants are metered into the chamber, then ignited. Liquid propellants have been studied because they produce lighter molecular weight combustion products, which leads to higher muzzle velocities, and because bulk liquids can be stored more compactly than shells, and require less handling equipment to load. Metering the propellant flow into the barrel lowers the peak pressure relative to a detonation, leading to lighter barrels.

Status:

Variations:

References:

13 Gaseous Charge

Alternate Names:

Type:

Description: A mixture of fuel and oxidizer is introduced into a chamber, and then ignited. This is similar to how automobile engines function. This method covers projectiles directly driven by the resulting high temperature gas. Indirect drive guns are covered under **Light Gas Guns** on the next page. The igniter is a separate device in the basic concept. An alternate version called a **Scramjet Gun** uses the projectile itself as a traveling igniter. It must be moving supersonic to keep the ignition front from getting ahead of it, so it needs a gas injector at the start to get it moving fast enough.

Status: Research prototypes. Methane/air mix was used as the driver for the Livermore 2 stage gas gun. The combustion drives a 1 ton piston, which in turn compresses hydrogen working gas. Research on the scramjet gun was being performed at the University of Washington under Prof. Adam Bruckner. Research gun was located in the basement of a building there.

Variations:

- **13a Scramjet Gun (Ram Accelerator)** - Fuel/oxidizer mixture present in barrel is burned as projectile travels up barrel. If projectile shape resembles two cones base to base, as in an inside-out scramjet, the gas is compressed between the projectile body and barrel wall. The combustion occurs behind the point of peak compression, and produces more pressure on the aft body than the compression on the fore-body. This pressure difference provides a net force accelerating the projectile.

One attraction of this concept is that a high acceleration launch can occur without the need for the projectile to use on-board propellants. If the projectile has a inlet/nozzle shape (hollow in the middle) it might continue accelerating in the atmosphere by injecting fuel into the air-only incoming flow, extending the performance beyond what a gun alone can do. Another attraction of this concept is the simplicity of the launcher, which is a simple tube capable of withstanding the internal pressure generated during combustion.

References:

- A. Hertzberg, A.P. Bruckner, and D.W. Bogdanoff, "The Ram Accelerator: A New Chemical Method of Accelerating Projectiles to Ultrahigh Velocities" , AIAA Journal, Vol. 26, No. 2, February 1988. (The original scramjet gun paper)
- P. Kaloupis and A.P. Bruckner, "The Ram Accelerator: A Chemically Driven Mass Launcher" , AIAA Paper 88-2968, AIAA/ASME/SAE/ASEE 24th Joint Propulsion Conference, July 11-13, 1988, Boston, MA. (Applications to surface-to-orbit launching)
- Breck W. Henderson, "Ram Accelerator Demonstrates Potential for Hypervelocity Research, Light Launch," , Aviation Week & Space Technology, September 30, 1991, pp. 50-51.
- J.W. Humphreys and T.H. Sobota, "Beyond Rockets: the Scramaccelerator" , Aerospace America , Vol. 29, June, 1991, pp. 18-21.

14 Rocket Fed Gun

Alternate Names:'

Type:

Description: A rocket engine is mounted at the chamber end of a gun to produce hot gas to accelerate projectile. In a conventional gun, all the gas is formed at once as the charge goes off. In this concept the gas is produced by a rocket type engine and fills the barrel with gas as the projectile runs down it. Compared to a conventional gun, the peak pressure is lower, so the barrel is lighter

Status:

Variations:

References:

NOTE: Guns and Accelerators is continued on the next page.

Section 2.2 - Guns and Accelerators (page 2)

Note: Continued from sections A and B at Guns and Accelerators

C. Light Gas Guns

Light gas guns are designed to reach higher muzzle velocities than combustion guns. These types of guns are also called **hypervelocity guns** since the projectile travels at more than Mach 5 (1500 m/s at sea level). In aerodynamics, velocities above Mach 5 are are called hypersonic. They reach higher velocities by using hot hydrogen (or sometimes helium) as the working gas. These have a lower molecular weight, and therefore a higher speed of sound than the combustion products of smokeless powder, liquid, or gaseous fuels.

Guns are strongly limited by the speed of sound of the gas they use, since pressure waves travel at the speed of sound. Therefore any gas at the back of the barrel no longer contributes to pushing a projectile once it reaches the local speed of sound, because the pressure waves cannot keep up. The gases near the projectile and moving at the same speed can still provide pressure, though. So the efficiency of the gun falls significantly, but does not stop entirely as the projectile goes supersonic. Light gases like Hydrogen do not generate high pressures and temperatures by themselves, as do combustion byproducts in standard guns. Therefore some external means are required to produce the hot gas. The various methods are listed in the following sections.

Scaling - Light gas guns have been used for hypersonic research for about 40 years. Examples are testing heat shield materials, and meteoroid impact damage. These do not require large projectiles, so a light gas gun large enough for practical use in space transport has not been built yet. There is no reason to think larger guns will not work. Higher muzzle velocities than needed for space launch have already been demonstrated in research guns. It is more a matter of designing them for low cost operation rather than research, and simple scaling of the parts. Large guns will in fact be more efficient than research guns:

- The region close to the barrel wall is called the **boundary layer**. This is where heat is lost to the cooler barrel, and friction is created between the moving gas and stationary wall. The boundary layer stays the same thickness regardless of gun size, so represents a smaller fraction in larger guns.
- Atmospheric drag is proportional to the area of the projectile, while it's mass is proportional to area x length. The mass thus grows faster than drag, and deceleration due to drag goes down with size. So drag induced velocity loss becomes a smaller factor for larger guns.

Larger guns also tend to have lower acceleration of the projectile. The muzzle velocity of this type of gun is mostly set by the speed of sound of Hydrogen, and the practicality of flying through the atmosphere at high speeds. So the muzzle velocity will not be greatly different in larger guns. The barrel length goes up with gun size, so the acceleration is lower to reach the same final velocity. That means the projectile structure can be lighter, and a wider range of cargo can be carried. In the limit of very large guns, it would be possible to carry humans, but such large guns will not be built for some time, and might be overtaken by better alternatives.

Comparison to other Launchers - Compared to rockets, weight is not an issue for a terrestrial mounted gun. Thus their parts can be closer to industrial grade than aerospace grade, and should be relatively inexpensive. Gun barrels are also simple compared to electromagnetic coil type accelerators providing the same force. Thus they should have a lower construction cost. Electromagnetic devices have a higher theoretical efficiency and maximum velocity than a gas gun, so the choice will depend on construction vs operations costs of real designs. Electromagnetic guns have a shorter history, so their capability and cost at large scale are more uncertain.

Gun/Rocket Optimization - When launching from Earth, the projectile encounters the atmosphere at high velocity, and drag tends to scale as velocity squared. Even with light gas guns, it is difficult to reach full orbit velocity (~8 km/s), and even at that velocity the orbit will intersect the ground, because that is where it started from. Thus any gun system needs at least some propulsion to raise the low part of the orbit. When optimized for cost and efficiency, the gun velocity generally ends up being roughly half of orbit velocity, with the balance provided by a rocket on the projectile or some other method. A rocket that only has to provide a fraction of orbit velocity will be much smaller and less expensive, and repetitive smaller launches lower the size even more. For example, daily gun launches compared to about six conventional rocket launches per year gives a factor of 60, and approximately 4 times lower rocket to payload mass ratio gives a total reduction of 240 in rocket size. The main cost advantage of hypervelocity guns thus comes from replacing the large and expensive conventional rocket with a much smaller one plus a reusable and relatively cheap gun.

Location - Due to their size, light gas guns will likely be built on a mountain with the correct slope, pointing east. This minimizes construction cost and reduces air drag from the higher starting altitude. An ocean platform for the gun could be aimed, but has the drawbacks of more atmosphere to fly through at sea level than from a mountain top, and dealing with salt water corrosion. When choosing a mountain, equatorial ones allow meeting a space station every orbit (approx 90 minutes), while other latitudes only allow meeting once per day, when the launch site rotates under the orbit plane.

Evacuated Barrel - All large guns will operate more efficiently if the projectile does not need to push a column of air up the barrel. So some method of sealing the end and pumping out most of the air is generally desired. In smaller guns something as simple as a plastic film has been used, and the projectile just punches through it. In larger guns a flap or flaps held in place by the pressure difference of the barrel vs outside air can be forced open as the residual air in the barrel gets compressed by the projectile. This sort of automatic opening is preferred to some mechanical valve or door. A failure of the mechanical device would lead to a spectacular collision with the projectile.

Muzzle Design - It is worth considering a **silencer** type device at the muzzle for several reasons. In addition to sound reduction, it can capture the Hydrogen gas to be recycled. Otherwise escaping hot Hydrogen will immediately burn in air, producing a muzzle flash. When the projectile goes from high acceleration in the barrel to deceleration in air, if the transition is too sudden that might cause damage to the projectile or cargo. A silencer can bleed off some of the Hydrogen, giving a smoother transition. The residual air in the barrel will also provide some degree of shock buffer as it piles up ahead of the projectile. Other ways to lower the shock are to put gas nozzles around the muzzle to add a stream of gas as a transition region past the muzzle, or to shape the barrel as a flaring con at the end to taper off the acceleration. Finally, the on-board rocket on the projectile can be used to counter the drag deceleration.

Projectile Features - Filling the area behind the projectile with gas, known as **base bleed**, has been used previously to reduce drag and extend the range of artillery. The velocity of light gas gun projectiles is high enough that shaping the back end to reduce drag may not be practical, and so base bleed is an option. This can be combined with the on-board rocket engine by running it at low thrust merely to fill the trailing area, or at full thrust to provide acceleration even while clearing the atmosphere. The right answer will require detailed design and analysis.

Projectile Recovery - To lower cost, you would like to use your cargo projectiles multiple times. Fortunately, making them rugged enough to fire out of a gun tends to also make them rugged enough to survive re-entry without a lot of extra work. They already need some heat shielding to survive going up through the atmosphere at high velocity when launched. You just need some more shielding to also handle re-entry. The same guidance system that enables them to reach their cargo destination also can guide it to a landing point. The terminal velocity of the empty projectile after re-entry should be low enough that the projectile can withstand landing without any landing system. It was fired fully loaded at high gees out of the gun, after all. At most it might need deployable flaps or fins or a small parachute to get the landing velocity low enough.

15 Pressure Tank Storage Gas Gun

Alternate Names:

Type:

Description: The gas is simply stored in a pressure vessel, then adiabatically expanded in a barrel, doing work against a projectile. Heating the gas raises issues of storage at high temperature, but not heating the gas leads to low performance. The storage tank needs to be large relative to the barrel volume, otherwise the pressure drop during firing leads to lower performance. This leads to higher cost than other versions. These drawbacks lead to other versions of a light gas gun to be preferred over this simple one.

Status:

Variations:

References:

- Taylor, R. A. "A Space Debris Simulation Facility for Spacecraft Materials Evaluation", SAPE Quarterly , v 18 no 2 pp 28-34, 1987.

16 Underwater Storage Gas Gun

Alternate Names: Quicklaunch

Type:

Description: In a gas gun on land the amount of structural material in the gun is governed by the tensile strength of the barrel and chamber. In an underwater gun, an evacuated barrel is under compression by water pressure. The gas pressure in the gun can now be the external water pressure plus the pressure the barrel wall can withstand in tension, which is up to twice as high as the land version

Other features of an underwater gun are the ability to store gas with very little pressure containment (the storage tank can be in equilibrium with the surrounding water), and the ability to point the gun in different directions and elevations.

The underwater gas gun consists of a gas storage chamber at some depth in a fluid, in this case the ocean, a long barrel connected to a chamber at one end and held at the surface by a floating platform at the other end, plus some supporting equipment.

Example Design In one version of this type of gun, the chamber is a made of structural material such as steel. An inlet pipe allows filling of the chamber with a compressed gas. A valve is mounted on the inlet pipe. An outlet pipe of larger diameter than the inlet pipe connects to the gun barrel. An outlet valve is mounted on the outlet pipe. This valve may be divided into two parts: a fast opening and closing part, and a tight sealing part. The interior of the chamber is lined with insulation. The inner surface of the insulation is covered by a refractory liner, such as tungsten. An electrical lead is connected to a heating element inside the chamber

An inert gas such as argon fills the insulation. The inert gas protects the chamber structure from exposure to hot hydrogen, and has a lower thermal conductivity. An inert gas fill/drain line is connected to the volume between the chamber wall and the liner. A pressure actuated relief valve connects the chamber with a volume of cold gas. This cold gas is surrounded by a flexible membrane such as rubber coated fiberglass cloth.

In operation, the gas inside the chamber, the inert gas, and the water outside the chamber are all at substantially the same pressure. Thus the outer structural wall does not have to withstand large pressure differences from inside to outside. One part of the chamber wall is movable, as in a sliding piston, to allow variation in the chamber volume. The gas in the chamber is preferably hot, so as to provide the highest muzzle velocity for the gun. When the gun is operated, this gas is released into the gun barrel. In order to preserv the small pressure difference across the wall of the chamber, either the chamber volume must decrease or gas from an adjacent cold gas bladder must replace the hot gas as it is expelled. This arrangement prevents ocean water from contacting the chamber walls or hot gas. In the case of the sliding piston, the membrane collapses, with the gas formerly within it moving in behind the piston. In the alternate case, the membrane also collapses, with the gas formerly within it moving through a valve into the chamber

The chamber has an exit valve which leads to the gun barrel. It also has gas supply lines feeding the interior of the chamber and the volume between the chamber walls. These lines are connected to regulators which maintain nearly equal gas pressures, which in turn are nearly equal to the ocean pressure. This allows the chamber to be moved to the surface for maintenance, and to be placed at different depths for providing different firing pressures or different gun elevations.

The muzzle of the gun is at the ocean surface, so elevation of the gun can be achieved by changing the depth of the chamber end. Since the gun as a whole is floating in the ocean, it can be pointed in any direction. Some means for heating the gas stored in the chamber is needed, such as an electric resistance heater. At the muzzle end of the gun, a tube surrounds the barrel, with a substantial volume in between the two. There are passages through the wall of the barrel that allow the gas to diffuse into the tube rather than out the end of the gun, thus conserving the gas.

At the muzzle of the gun is a valve which can rapidly open, and an ejector pump which prevents air from entering the barrel. In operation, the ejector pump starts before the gun is fired, with the valve shut. The valve is opened, then the gun is fired. In this way, the projectile encounters only near vacuum within the barrel, followed by air

Status: Hypersonic guns have been in operation for about 40 years as research devices, with relatively small (5 kg or less) projectiles. Large or ocean-going guns have not been built.

Variations:

References:

17 Particle Bed Heated Gas Gun

Alternate Names:

Type:

Description: Hot hydrogen gives good performance for a gun, but is hard to store. In this method it is generated when needed by flowing the Hydrogen through a chamber which contains refractory oxide particles. The particles are heated slowly (roughly 1 hour time period) by some type of heater near the center of the chamber. This sets up a temperature gradient, so the exterior of the chamber is relatively cool, and can thus be made of ordinary steels. When the hydrogen flows through the chamber, the large surface area of the particles allows very high heat transfer rates - so the heat in the chamber can be extracted in a fraction of a second as the gun fires. The gas flows inward from the periphery to the center of the heat exchanger and then to the barrel.

- **Example of Gun Component Sizing:** - The following calculations are intended to show the steps for getting the initial estimates of component sizes. It is very far from being a complete engineering design. Rather, it is the starting point for the design process.

  - Set as a design goal to match the 1990's Livermore SHARP hypersonic gun (5 kg at 3 km/s). From kinetic energy formula (KE = 0.5*M*v^2) we get 22.5 MJ projectile energy.
  - If average barrel pressure is 20.7 MPa (3000 psi), chosen as a reasonable value for high pressure pipe, then a 10 cm diam, 5 kg projectile (matching SHARP parameters) will see 162,600 N force, or 32,470 m/s2. To reach 3 km/s requires 92.4 msec, and a barrel length of 139 m. Allowing for pressure drop during operation, friction, efficiency, thermal, and other losses, assume actual barrel will be 200 m long. That assumption would be later replaced by a simulation to get a better estimate.
  - From simple cylinder volume formula (V = pi * r^2 * h) the volume of the barrel will be 1.57 cubic meters. From the physical properties of Hydrogen as a gas, that volume would contain 4 kg of Hydrogen at 20 MPa and 2000K temperature. At the point the projectile leaves the barrel, the barrel is full of gas. At standard conditions (273K or 0C), the speed of sound for hydrogen is 1284 m/s. That goes up as the square root of the Kelvin temperature, so at 2000K it will be 3475 m/s, and so for this gun it will be slightly subsonic.
  - The specific heat of Hydrogen from 300 to 2000K is about 15 kJ/kg-K. Therefore by multiplying the known temperature rise (1700K), mass of H2 (4 kg) and specific heat, we need 102 MJ of energy for heating. To this we add the 22.5 MJ of projectile kinetic energy, which comes at the expense of the Hydrogen heat and pressure, for a total of 124.5 MJ.
  - To heat the gas during the launch, we don't want the aluminum oxide particles to drop more than 500K in operation. They otherwise would be cooling down too much to heat the last part

of the gas. At 1300 J/kg-K, need then need about 200 kg of aluminum oxide particles. That is a rough value which would be refined by a thermal analysis of the heat exchanger.
- The density of Aluminum Oxide is 3.9 g/cc as a solid. As grains in a heat exchanger with allowance for inlet and outlet volume and screens to keep the particles in place, assume the density is 1 g/cc. Therefore heat exchanger will be about 0.2 cubic meters in size.
- The storage tank for the unheated hydrogen can be found from the physical properties at room temperature. The initial pressure will have to be higher than 20 MPa for the gas to flow towards the barrel. Assume it starts at 1.5 times the average barrel pressure and finishes at 0.5 times. Thus 2/3 of the gas goes into the barrel, and the initial gas to be stored is then 6 kg at 31 MPa. Under those conditions the density of Hydrogen is 25.8 kg/m^3, so the tank volume is 0.23 cubic meters. Very quickly releasing the gas from the storage tank turns it into a rocket engine, so a sturdy support is needed to keep the tank in place.
- A fast (compared to the < 0.1 sec firing time of the gun) valve will be needed to open the tank and let the gas flow through the heat exchanger and barrel. Conceptually this can be like a car engine cylinder, where a spark plug sets off a detonation, which slides the valve piston from the closed to open position. When doing the design, you would first review existing valve hardware to see if a suitable one exists. If not, then a custom fast-acting valve would need to be designed.

Status: A small research gun of this type has been built at Brookhaven Natl. Lab.

Variations:

References:

18 Particle Bed Reactor Heated Gas Gun

Alternate Names:

Type:

Description: Hot gas is generated by flowing through a particle bed type nuclear reactor. Gas expands against projectile, accelerating it. Light gas guns have been operated to above orbital velocity, and 1 kg projectiles have been accelerated to over half orbital velocity This type of gun rapidly becomes less efficient above the speed of sound of the gas. As a consequence the working fluid is usually hot hydrogen. Conventional gas guns have used powder charge driven pistons to compress and heat the gas. This is not expected to be practical on the scale needed to launch useful payloads to orbit. One way to heat the gas is to pass it through a small particle bed nuclear reactor. This type of reactor produces a great deal of heat in a small volume, since the small particles of nuclear fuel have a large surface/volume ratio and can efficiently transfer the heat to working fluid. This gives the benefits of nuclear power for space launch, without the drawbacks of a flying reactor.

The particles are retained in the reactor by spinning the bed at high velocity, and the gas flow cools the external structure. The particles of nuclear fuel do not need structural strength, so can go to higher temperatures than a solid core reactor, leading to higher performance. The improvement, however, is not great over refractory metal heat exchangers, so the cost and political issues for a nuclear device probably outweighs the benefits.

Status:

Variations:

References:

19 Electrically Heated Gun

Alternate Names:

Type:

Description: The working fluid is heated by electric discharge or microwave induction, then pushes against the projectile in the barrel. The limiting factor for a light gas gun is the speed of sound in the fluid. One way to heat the fluid to much higher temperature is an electric discharge. If a high enough temperature is reached, the fluid becomes a plasma. Either magnetic fields or a sheath of cooler gas may be used to keep the plasma away from the walls and prevent damage. The challenge for this method is to deliver hundreds of MegaJoules for even a very small gun (by space launch standards) in a short period of time. Magnetic fields are not strong enough to contain the plasma at these pressures, so it will lose energy rapidly to physical chamber walls. Therefore the heating must be rapid, which in turn requires a large power source. For versions large enough to deliver payloads to orbit the power source needs to deliver 2 GigaJoules or more, and the power source ends up dominating the total gun cost.

Status:

Variations:

References:

20 Nuclear Charge Heated Gun

Alternate Names: Nuclear Cannon

Type:

Description: As in the other light gas guns, the objective is to use hot gas to accelerate a projectile. In this version the heating is from a small atomic bomb in a large underground chamber filled with the working gas. This concept only makes sense in a situation where very large payloads need to be launched, due to the large minimum energy to make any nuclear device function. A large barrel leads off the chamber upward at an angle. A crossbar is set into the barrel near the chamber, and the projectile is attached to the crossbar with a bolt that is designed to fail at a pre-determined stress. This restrains the projectile until the operating pressure is reached. A small atomic bomb is suspended in the chamber and detonated to create lots of very hot hydrogen in a very short time.

There are several obvious issues with this concept. Even a small atomic bomb delivers too much energy in too short a time to easily manage. The challenge is to keep from destroying the gun itself. Another issue is radiation, which can harm the cargo, and get ejected out the barrel as fallout.Finally are the political restrictions on using atomic bombs of any purpose.

Status:

Variations:

References:

21 Combustion Driven Two Stage Gun

Alternate Names:

Type:

Description: This is a type of two-stage gas gun. A cylindrical chamber contains a piston. On the back side of the piston high pressure gas is generated by combustion. This can be gunpowder or a fuel-air mixture. On the front side of the piston is the working gas, which is usually Hydrogen. The Hydrogen is compressed and heated until a valve or seal is opened. Then the working gas accelerates the projectile. The compression can result in theoretically unlimited temperature and pressure. In practice it is limited by the pressure capacity of the chamber structure, but this can be made very high. Therefore muzzle velocities have reached above Earth orbit velocity (8 km/s) for small projectiles. The drawback is the chamber must be several times the barrel volume to contain both the combustion driver gas and Hydrogen, which increases the cost for large versions.

Status: This type of light gas gun is the most common that has been built. They were first constructed in the 1960's or earlier. The largest gun of this type was the Lawrence Livermore Laboratory SHARP gun, which was used to test scramjet components in the early 1990's. It had a 10 cm x 45 m barrel and a 30 cm x 100 m long chamber.

Variations:

References:

- Light Gas Gun (Wikipedia article)
- Aviation Week & Space Technology, July 23, 1990.
- "World's Largest Light Gas Gun Nears Completion at Livermore." Aviation Week & Space Technology, August 10,1992.

22 Gravity Driven Piston

Alternate Names:

Type:

Description: In this method a sliding or falling mass is used to compress gas in a chamber. The gas is then expanded in a barrel. As in all light gas guns, hot Hydrogen gas is the best working fluid, and the question is how to generate it. In the previous method a piston is driven by combustion, much like an internal combustion automobile engine. Here a good part of the high pressure chamber is dispensed with by using gravity to drive a massive weight. If the gun is built on the side of a mountain, as is usually desired, the energy for launch is stored as potential energy in the weight. The falling weight rides on an air or lubricated bearing and slides down the mountain to the chamber. The piston seals the back end of the chamber and is held in place. The falling mass is designed smaller than the piston, so it does not need an accurate fit at the high velocity it impacts. The chamber then leads to a barrel containing the projectile, which accelerates upward.

One advantage of this method is the falling mass is not stopped instantly, but rather can continue compressing the gas gradually as the gun fires. The system of falling mass + working gas + projectile can be designed as a tuned coupled spring system to maximize energy transfer to the projectile. Another is the relatively low cost, since the falling mass operates in open air, and can be made from low strength materials such as concrete. A third is the mass can be raised as slowly as needed, so large power supplies are not needed. Lastly this method can be added as a "supercharger" to a particle bed type heater to increase performance. The gas is first heated to the practical limit of flowing through a heat exchanger bed, and then the weight driven piston provides additional compression heating. The question then becomes whether the additional performance is needed and justifies the added complexity of a dual heating and compression system.

Status: As of yet this is only a concept.

Variations:

References:

D. Electromagnetic Accelerators

This group uses magnetic fields to provide a force between the accelerator and a projectile. Electric accelerators of moderate length require high peak power for a short period of time. Hence inexpensive energy storage is very important for these concepts. Two places to look for inexpensive energy storage are (1) Magnetic fusion experiments, which also need short duration high power supplies, and (2) Inductive energy stores. The latter falls into subcategories of cooled normal conductors, and superconductors. Alternately the peak power can be lowered by making the device very long, but that then becomes a construction issue to build a device many kilometers in size.

23 Railgun

Alternate Names: Electromagnetic Gun

Type:

Description: In this method, very high electric current supplied by massive rails is shorted through a plasma arc induced at the back end of the projectile. The plasma is accelerated by the Lorentz force interaction with the magnetic field produced by current. The plasma then pushes projectile along the rails. Given a sufficiently large power supply, it can be considered for Earth launch systems at lower accelerations than those proposed for weapon systems.

Status: This device was under intensive development for the Strategic Defense Initiative. A large gun was built at Eglin AFB in Florida and used a bank of thousands of car batteries wired in parallel as a power supply. Prototype railguns achieved high velocities, but the high currents and plasma produced rail erosion.

Variations:

References:

- Robinson, C. A. "Defense Department Developing Orbital Guns", Aviation Week and Space Technology, v 121 no 12 pp 69-70, 1984.
- Bauer, D. P. et al "Application of Electromagnetic Accelerators to Space Propulsion" IEEE Trans. Magnetics vol MAG-18 no 1 pp 170-5, Jan. 1982.

24 Coilgun

Alternate Names: Mass Driver

Type:

Description: In this method a series of coils forming the gun react with coil(s) on projectile magnetically, producing linear thrust. Popularly known as a **Mass Driver**, it functions on the same principles as an electric motor. The concept was originally developed in connection with launching Lunar materials for space manufacturing. The low orbital velocity and vacuum on the Moon made it feasible. Accelerator designs with high efficiency (>90%) and high muzzle velocities (>8 km/s) have been proposed for use on Earth. This potentially leads to a transportation system whose operating costs consist mostly of electricity. Because of the need for large amounts of power to drive the device, the cost of that power supply tends to dominate overall system cost, so an electromagnetic accelerator makes sense for very high volume launch.

Like other launchers that generate high velocity near the Earth's surface, some method of dealing with atmospheric drag and heating is needed. Options include heat shielding on the projectile, a vacuum tunnel, or installing the launcher on high enough towers to avoid the atmosphere.

Status: Linear electric motors and magnetic levitation trains operate on the same principle as a coilgun. A coilgun is just a higher performance version. Prototype coilguns were built around 1980 and reached 1800 gravities acceleration, which is more than sufficient.

Variations:

24a Quenchgun - The energy is stored in superconducting coils making up the gun. The circulating current is **quenched** ahead of the projectile either by heating the coil above the transition temperature of the superconductor, or by raising the field using a coil on the projectile. In either case the current stops flowing in the gun coil. Since the coils behind the projectile are off, while those ahead are still on, the net force will accelerate the projectile.

References:

- Nagatomo, Makoto; Kyotani, Yoshihiro "Feasibility Study on Linear-Motor-Assisted Take-Off (LMATO) Of Winged Launch Vehicle", Acta Astronautica, v 15 no 11 pp 851-857, 1987.
- Kolm, H.; Mongeau, P. "Alternative Launching Medium", IEEE Spectrum, v 19 no 4 pp 30-36, 1982.
- Kolm, H. "An Electromagnetic 'Slingshot' for Space Propulsion", Spaceworld pp 9-14, Feb. 1978.

Section 2.3 - Combustion Engines

A. Air-Breathing Engines

The next group of methods normally involve using a planet's (usually the Earth's) atmosphere as a supply of oxygen to support combustion with a fuel carried on the vehicle. They differ in the details of how the incoming air flow and combustion is managed. It should be noted that some vehicle concepts, such as the National Aerospaceplane (NASP) of the 1990's, or the current British Skylon would integrate more than one method in a single engine. This is referred to as a **Combined Cycle Engine** The same general engine concepts could be used in a reducing atmosphere, such as Hydrogen or Methane, with Oxygen as the carried fuel, or in a sufficiently powerful nuclear engine with any atmosphere. In the latter case the nuclear engine is used to drive a compressor or heat the incoming gas flow.

Fundamental to understanding the operation of air-breathing engines is the concept of mass conservation. The mass of incoming air at the **Inlet** does not change in total amount going through the engine, although pressure and temperature vary. The exception is the point where fuel is added, or in some types, where condensed air flow is removed. Heat can be added or removed by combustion or heat exchangers, but otherwise the mass flow stays the same. From that constant mass flow, changes in the other conditions of the gas flow can be calculated, and performance derived.

In practice, the flow through an engine can involve many fan blades and turbulence, so that computer simulations are not fully accurate. Real engines also operate at extremes of temperature and stress. Therefore engine development typically includes extensive testing on static test stands, in wind tunnels, and in flight attached to aircraft, to determine actual performance and durability

25 Fanjet

Alternate Names:

Type: Gas Flow by Fuel-Atmosphere Combustion

Description: The fanjet is the standard type of jet engine found on passenger aircraft and military aircraft. The original form of the engine, the turbojet, has a series of turbine compressor stages to compress the incoming air flow. This is followed by a combustor where fuel is added and burned, creating a hot gas. The gas is then expanded through a turbine which is connected by a shaft to the compressor. The expanded gas emerges at high velocity from the back of the engine.

The modern fanjet adds a fan which is also driven by the turbine. All of the airflow goes through the fan, but only a part goes into the compressor. The air which does not go into the compressor is said to have 'bypassed' the compressor. The 'bypass ratio' is the ratio of bypass air to combustor air. Generally higher bypass ratio engines are more fuel efficient, in units of thrust divided by fuel consumption rate, because they increase the mass flowing through the engine. In general, engines that operate at higher speeds are designed with lower bypass ratios. This is due to more combustion needed to increase thrust against the higher inlet drag from the higher velocity.

Typical modern performance values are engine thrust-to weight ratios (T/W) of 6:1 for large subsonic engines, trending towards about 10:1 for high performance military jets. Fuel efficiency is measured in units of thrust divided by mass flow rate. In English units this is pounds divided by pounds per second, or just seconds, and is termed 'specific impulse'. In SI units this is Newtons per kilogram per second, which has the units of meters per second. In some propulsion systems, such as chemical rockets, the SI unit corresponds to the actual exhaust jet velocity. In the case of air-breathing propulsion it is not, the velocity result is just an indicator of engine efficiency. In English units the performance of subsonic engines is about 10,000 seconds, trending to about 7000 seconds for supersonic military engines. Fanjets and turbojets operate up to about 3.5 times the speed of sound (M=3.5). Extra thrust can be generated with an **Afterburner**, which burns more fuel after the turbine stage of the engine. This comes at the cost of lower fuel efficiency.

Although the maximum altitude and velocity of a fanjet is limited compared to Earth orbit velocity, the effect on payload can be much larger. This is because compared to a conventional rocket, it can avoid most of the drag, pressure, and gravity losses near the ground. Conventional rocket payload is typically a small fraction of total mass, so small reductions in losses can produce large relative increases in payload.

Status: In common use on aircraft for aircraft propulsion. For space launch, the B-52 bomber and L-1011 aircraft have been used to carry the Pegasus three stage solid rocket to 35,000 ft altitude. The B-52 uses 8 fanjet type engines for propulsion. The Stratolaunch system is under development, using parts from two 747 aircraft with 6 high bypass fanjet engines, and carrying a large rocket. Numerous paper studies have been made of using aircraft as carriers for rocket stages.

Variations:

- **Carrier Aircraft** - A conventional jet aircraft is used to carry a separate rocket to an altitude of roughly 9000 meters and velocity of 240 m/s, after which the rocket ignites and finishes reaching orbit.

- **Booster Jets** - A set of military fighter engines are attached to a rocket as separate strap-on boosters, or a connected **booster ring**. They can loft a rocket to about 15 km altitude and 480 m/s velocity, after which they either parachute land or do a powered vertical landing.

References:

26 Turbo-Ramjet

Alternate Names:

Type: Gas Flow by Fuel-Atmosphere Combustion

Description: In this method a multi-stage fan compresses the incoming air stream, which is then mixed with fuel, burned and exhausted. The compressor is driven by a gas generator/turbine. In a fanjet, the incoming air is compressed and heated by the compressor stages, then mixed with fuel and run through the turbine stages. At higher velocities the air gets hotter in compression since it has a higher incoming kinetic energy. This leads to a higher turbine temperature. Eventually a turbine temperature limit is reached based on the material used, which sets a limit to the speed of the engine. In the turbo-ramjet the compressor is driven by a gas generator/turbine set instead, which use on-board propellant for their operation. Since the gas generator is independent of the flight speed, it can operate over a wider range of Mach numbers than the fanjet (to Mach 6 vs. to Mach 3).

Status:

Variations:

References:

27 Ramjet

Alternate Names:

Type:

Description: In this method, the incoming air stream is decelerated to subsonic velocity relative to the engine via a shaped inlet, mixed with fuel, then accelerated again out an exit nozzle. Conceptually it is the simplest form of jet engine, because it has no fans or turbines. The incoming air is moving at the vehicle velocity entering the engine. After burning the fuel, the air is hotter and can expand to a higher velocity out the nozzle. This sets up a pressure difference that leaves a net thrust. Ramjets cannot operate at zero speed, but they can reach somewhat higher limits than an engine with rotating machinery (range Mach 0.5 to about Mach 8).

Status:

Variations:

- **Air Augmented Rocket-** This is a form of combined cycle engine. Since ramjets do not function at zero speed, an internal rocket chamber in the engine is used for initial thrust. By entraining air flow, the thrust level can be augmented at lower speeds. Once ramjet speeds are reached, the engine functions in ramjet mode, using the rocket chamber as a fuel injector. At the upper end of ramjet function, the engine transitions back to pure rocket mode.

References:

- Wikipedia article: **Air-augmented Rocket**

28 Scramjet

Alternate Names:

Type:

Description: This is similar to how a ramjet functions. The incoming air stream is compressed by shock waves, mixed with fuel, and expanded against the engine or vehicle body. The difference is the airstream remains supersonic relative to the vehicle. The ramjet requirement to slow the airstream to subsonic speed becomes inefficient at higher velocities. Even though the gas is moving supersonically relative to the vehicle, the sidewise expansion can act on the vehicle if the slope of the nozzle is low enough. Thus the vehicle can fly faster than the exhaust gas moves. Scramjets may provide useful thrust up to about Mach 15, or 60% of Earth orbital speed.

The very high velocities lead to extreme heating of vehicle parts. High compression and expansion efficiency is needed to get a positive net thrust, since the energy added by the fuel becomes small relative to the kinetic energy of the air flow. Thus development of working scramjets has proved difficult. Scramjets also do not function at zero velocity, so some other method is needed to get to their starting point. Therefore complete vehicles will need a combined engine system.

Status: Scramjet engine components and small scale versions have been tested with mixed success.

Variations:

References:

29 Inverted Scramjet

Alternate Names: Buoyant Scramjet

Type:

Description: Ramjet and scramjet vehicles would prefer Hydrogen as a fuel because it gives higher performance, and can be used for cooling vehicle parts heated by the high velocity in air. Unfortunately Hydrogen is also very low density, which leads to relatively heavy vehicle structure due to large tanks. This method inverts the problem by using a series of balloons or a lightweight pipe supported in the atmosphere. They contain Hydrogen, and the vehicle carries Oxygen in its tanks. Oxygen is about 16 times denser than Hydrogen, so the tank size is much reduced.

Status: This is only a concept at present.

Variations:

References:

30 Laser-Thermal Jet

Alternate Names:

Type:

Description: In this method a laser beam is focused on and absorbed by a heat exchanger on the vehicle, or creates a laser-sustained plasma. The hot gas is then exhausted for thrust. By not requiring fuel, it is potentially efficient. The drawback is it requires a very powerful laser to be feasible even for small vehicles. Powerful lasers are currently expensive. Another limitation is the distance over which the beam can maintain focus.

Status: Minimal experiments have been done in a laboratory.

Variations:

References:

- Myrabo, L. N. "Concept for Light-Powered Flight", AIAA paper number 82-1214 presented at AIAA/SAE/ASME 18th Joint Propulsion Conference, Cleveland, Ohio, 21-23 June 1982.

B. Internally Fueled Engines (Rockets)

Conventional rockets function by expelling a gas at high velocity in a desired direction. By conservation of momentum (a physical law), the remainder of the vehicle will gain velocity in the opposite direction. Rockets have been the principal method of space transport to date because they only require internally stored fuel, and can thus operate in a vacuum. Since they propel the vehicle, the fuels are also called **Propellants**. Rockets can be sorted into types by how many propellants they have and what physical form the propellants are stored in.The former category includes single fuel **Monopropellant**, two fuel **Bipropellant**, and the rarely used three fuel **Tripropellant** types. The category includes gaseous, solid, liquid, and hybrid - part solid and part liquid.

Thrust is mass flow rate times exit or **Exhaust Velocity**. To get the most use from a finite amount of fuel, you want to minimize flow rate and maximize exit velocity. Therefore the gas should be as hot as possible and have low molecular weight. That in turn drives the choice of chemical reaction and fuels to use. There are numerous fuel combinations that are possible, but a relatively few that have a combination of high energy and other desirable characteristics such as density, safety, and low corrosion. There are several ways to get the gas hot: catalytic decomposition, combustion, or external heating. The first two are grouped under **chemical rockets**, and the latter are categorized by how the gas gets heated. Chemical rockets generally use a combination of combustion chamber and expansion nozzle, as that is a very efficient way to direct the gas flow at high velocity. Rocket engines will function with a surrounding atmosphere, but that impedes the gas flow, so they generate lower thrust. The lost thrust can be approximated by the local exterior pressure times the area of the nozzle exit.

Rockets are generally less efficient than air-breathing engines in terms of momentum per fuel mass, since the latter can use oxygen from the air as part of the fuel. This increases the combustion energy per carried fuel mass. Air breathing engines also increase the mass flowing through the engine via the Nitrogen component of air, and additional un-combusted air flow using engine-driven fans.

Both rockets and air-breathing engines involve similar design principles, as they both use combustion and hot gas flow to get reaction forces.

History - The earliest reference to using expelled mass for reaction force is the Greek Archytus around 400 BC. [1] This used external heating to generate steam. The first known use of internal chemical energy is in 1232 AD in China by **fire arrows** using gunpowder as fuel. The idea may have been transmitted to Europe by the Mongols, where experiments and use for fireworks occurred in the 13-15th centuries. Experiments began in the 18th century as a transport method rather than explosive device, although military use continued.Notably, rockets used in a battle in 1812 were recorded in the US national anthem.

The use of rockets to reach space was proposed by Konstantin Tsiolkovsky in 1898. Robert Goddard built experimental solid and liquid fuel rockets starting in 1915. A book published by Hermann Oberth in 1923 influenced the formation of rocket societies where groups pursued their development. [2] The German government pursued the development of rockets to deliver explosives by sub-orbital trajectories from 1937 to 1945. Subsequently the scientists and their hardware and data went to the United States and the USSR, where they helped develop sub-orbital ballistic missiles and orbital transport, both first flying in 1957. The first orbital rocket were essentially identical to ballistic missiles, but have since diverged. By 1963 Liquid Hydrogen/Liquid Oxygen propellants were in use, which is still the highest energy fuel mix commonly used. Liquid and solid chemical rockets are by far the most common space transport method, and are now built in a number of countries by both governments and private companies.

Design - The non-fuel mass of a rocket stage can be grouped into engines, tanks, and "other". Rocket engines can produce 400-1000 N/kg of engine mass in thrust, which is many times larger than the 9.8 N/kg force of gravity. For liftoff from the Earth, you want approximately 1.2-1.5 times the vehicle takeoff weight in thrust, so the engine component then has a mass of about 1.3-3% of the total vehicle. A large tank, such as the Shuttle External Tank, can weigh 4% of the fuel weight, but other tanks can range up to 10% of the fuel weight. 'Other' includes plumbing, parachutes, landing gear, heat shields, guidance systems, and such non-propulsion parts. It can range from 1% up to 10% of the total weight.

Older materials and components required 15% of the total liftoff mass to be vehicle other than fuel, assuming a single flight operating life. Modern materials require about 10% of the total mass for a many flight operating life. Structures tend to get heavier at the rate of 10% for each factor of 10 in life. This comes from the fatigue life of materials under load cycles (flights) as a function of stress. Lower stress and longer life requires thicker and heavier structural parts to distribute the load over. So a 100-use structure will be about 20% heavier than a one-use structure.

31 Solid Rocket

Alternate Names:

Type:

Description: A solid rocket consists of a high-strength casing, a nozzle, and a precast solid propellant grain which burns at a pre-designed rate. The grain is a mixture of materials containing both fuel and oxidizer, so combustion can proceed without any external action once it is ignited. Modern solid propellants have a formulation close to the following: About 15% by weight organic fuel, usually a type of rubber, about 20% by weight aluminum powder (which acts as a metallic fuel), and about 65% ammonium perchlorate ($NH3ClO4$), which is the oxidizer. About 1-2% epoxy is added to the powders to hold them together. The epoxy, being an organic material, is also part of the fuel. Solid propellants burn from the surface of the precast grain. Therefore the shape of the grain at ignition, and shape as it burns away determines the thrust level.

Advantages of solid rockets are short preparation time to launch and long term storage, compared to cryogenic fuels like liquid Oxygen. Disadvantages include relatively low exhaust velocity (2.6-3 km/s), and no easy way to turn it off or control it once ignited. They are often used as a booster first stage since the relatively dense fuel (1.35 g/cc) lowers the area of the vehicle. This is an advantage during the first two minutes of flight, where aerodynamic drag is important. When aluminum is used in the propellant, part of the end product is aluminum oxide, which is an excellent abrasive. Thus nozzle erosion is a significant effect that must be accounted for, in addition to the high temperatures.

Solid rockets are simpler in the sense of having few moving parts, but the entire motor casing that surrounds the fuel grain must withstand the operating pressure. In liquid rockets with fuel pumps, the propellant tanks see hydraulic and acceleration loads, which are typically lower, and only the pumps and combustion chamber see the full operating pressure.

Status: In common use for rocket stages, particularly the strap-on booster stage.

Variations:

References:

32 Hybrid Rocket

Alternate Names:

Type:

Description: The hybrid rocket consists of a solid fuel grain and a liquid oxidizer. One combination is rubber for the fuel and liquid oxygen for the oxidizer. The fuel is in the form of a hollow cylinder or perforated block. The oxidizer is sprayed onto the fuel and the material is ignited. By not being self-supporting in combustion, the fuel part can be treated as non- hazardous when being made and shipped. Only when on the launch pad and the oxidizer tank is filled is there a hazardous combination. With only a single liquid to handle, the hardware is relatively simple in design. Hybrid rockets are intermediate in performance compared to solid and full liquid engines.

Status:

Variations:

References:

33 Liquid Rocket

Alternate Names:

Type: Fuel/Oxidizer Combustion by Combustion Gas exhaust

Description: In a liquid rocket, the propellant ingredients are forced into a combustion chamber, where they burn, and which leads to a converging-diverging nozzle. The flow becomes sonic at the narrow part of the nozzle, then continues to accelerate in the diverging part of the nozzle, reaching about 1.5-2 times the speed of sound by converting temperature and pressure via expansion into a directed flow. A variety of propellant combinations have been used, including mono- bi-, and even tri-propellants. Monopropellants typically use catalytic decomposition for heating. The most common form of liquid rocket uses a separate fuel and oxidizer, which are mixed and burned in the combustion chamber. Many bipropellant mixtures are possible, but the highest energy-to-mass ratio mix in common use is from 1 part Hydrogen to 6 parts Oxygen. This produces mostly steam with a little leftover Hydrogen, which lowers the average molecular weight and thus increases the average molecule velocity. This propellant mix can reach about 4.7 km/s exhaust velocity under the best conditions.

Some propellant mixes will burn on contact, and so do not require an ignition source. These are called **Hypergolic** propellants. Some liquid propellants are liquid at room temperature and can be stored for long periods in a tank. These are referred to as **Storable**. Others, including Hydrogen, Methane, and Oxygen, are only liquid at very low temperatures. These are referred to as **Cryogenic**.

Status: This is the most common form of launch propulsion used to date to put things into Earth orbit.

Variations: There are multiple combinations of liquid engine types that are possible and have been used. They can be sorted by what propellant combination is used, how the propellants are delivered into the combustion chamber, and how the resulting hot gas is expanded out of it.

- **Variations by composition**- The following tables list some oxidizers and fuels, and combinations thereof. It is not an exhaustive list, and some ingredients have practical issues such as storage temperature, human toxicity corrosiveness, or chemical instability. Rocket propellants by their nature contain a lot of chemical energy and that energy can cause unintended reactions. Actual engine performance depends on factors such as chamber and exit pressures, so the table values should only be used for general comparisons. Kerosine is a mixture of compounds derived from petroleum, and Rocket Propellant 1 (RP-1) is a standardized type of Kerosine specified as a rocket fuel. Therefore it does not have an exact formula, and it is given as an approximate average value. It also does not have a well defined melting and boiling point, which is given as a range defined by distillation of its components.


A test firing of the Space Shuttle Main Engine, a Hydrogen/Oxygen propellant liquid engine.

Oxidizers

Chemical Name	Formula	Melting Point (K)	Boiling Point (K)	Density (kg/m^3)
Oxygen	O_2	54	90	1141
Hydrogen Peroxide	O_2H_2	273	423	1450
Fluorine	F_2	53.5	85	1505
Nitrogen Tetroxide	N_2O_4	262	294	1443
Chlorine Pentafluoride	ClF_5	170	260	1900

Fuels

Chemical Name	Formula	Melting Point (K)	Boiling Point (K)	Density (kg/m^3)
Hydrogen	H_2	14	20	70
Methane	CH_4	91	110	423
Propane	C_3H_8	85.5	231	582
Monomethylhydrazine (MMH)	$CH_3N_2H_3$	221	364	875
Kerosine (RP-1)	$\sim C_nH_{1.95n}$ n=~14	200	478-528	810

Propellant Combinations

Exhaust Velocities assume chamber pressure of 6.9 MPa and exit pressure of 0.1 MPa (Earth sea level) with optimum nozzle expansion.

Oxidizer	Fuel	Mass Ratio (O/F)	Exhaust Velocity (m/s)
Oxygen	Hydrogen	4.0	3820
Oxygen	Methane	3.0	3050
Oxygen	Kerosine	2.56	2942
Fluorine	Hydrogen	7.6	4021
Nitrogen Tetroxide	MMH	2.15	2834
Hydrogen peroxide (90%)	Kerosine	7.0	2912

- **Variations by fuel feed** - Some engine designs use a pump to feed the propellants into the combustion chamber, others use a pressurized tank. Large engines may use a combination of tank pressurization to prevent cavitation of the pump inlet and pumps to reach chamber inlet pressure. The fuel must enter the chamber at a higher pressure than the combustion pressure in a steady state engine. Non-steady state pulsed ignition engines are possible, but not generally used. Steady state engined deliver more continuous thrust. Pumps require a lot of power to operate, and generally use the same fuel as the rocket. In **Gas Generator** systems a portion of the fuel flow is used to create hot gas which drives a turbine to run the pump. The hot gas is then vented. In **Staged Combustion** systems, the hot gas is not completely burned, and is fed into the combustion chamber. This is more efficient but also more complex.

- **Variation by nozzle type** - Most rocket engines to date use a bell-shaped nozzle, where the gas flow is on the inside, surrounded by structure which directs the flow. An alternate design called an **Aerospike** nozzle inverts this arrangement, with the structure on the inside in the form of a wedge or cone, and the gas flow surrounding it. The outer edge of the gas flow is contained by the surrounding atmosphere. Since that automatically adjust for pressure differences, it is a type of altitude compensating nozzle. The benefit to compensation is that nozzle exit area in an atmosphere represents a thrust loss. A truncated cone can integrate better with some vehicle shapes, and the larger chamber and nozzle area can lower heat flux.

- **Variation by cooling type** - The high energy combustion in liquid rocket engines can exceed the melting point of most structural materials. Therefore methods to prevent this are needed in all but the smallest engines, which simply use high temperature alloys and radiate the heat away. In one method, the fuel is run through channels in the rocket engine walls to keep it from overheating. This also recovers some of the energy that would otherwise be lost. A method in recent development by **Orbitec** injects a counterflow vortex of one propellant ingredient along the inside of the engine walls, which is mixed with the other ingredient at the back and the hot gases then flow in the forward direction down the core of the chamber. The unburned ingredient protects the structure with a layer of cool gas.

References:

- Cooper, Larry P. "Status of Advanced Orbital Transfer Propulsion", Space Technology (Oxford), v 7 no 3 pp 205-16, 1987.
- Godai, Tomifumi "H-II Rocket: New Japanese Launch VehicleÊ in the 1990s", Endeavour , v 11 no 3 pp 116-21, 1987.
- Wilhite, A. W. "Advanced Rocket Propulsion Technology Assessment for Future Space Transportation", Journal of Spacecraft and Rockets, v 19 no 4 pp 314-19, 1982.

34 Gaseous Thruster

Alternate Names:

Type:

Description: In this method the propellant is introduced in gas form to the chamber. It may be a mono-propellant (a single gas) or a bi-propellant combination. Due to high tank mass this is usually used for small auxiliary thrusters. By using direct pressure from the tank to make the propellant flow, it can be very simple.

Status:

Variations:

References:

35 Mechanically Augmented Thruster

Alternate Names:

Type:

Description: The velocity of the exhaust gases are increased by placing the thrusters on the end of rotating arms. This can add 2-3 km/s to the exhaust velocity based on structural limits. It requires some external energy input to maintain the rotation of the arms, since the thrust opposes their rotation. Full electric thrusters generally have higher performance than this method, so they are preferred.

Status: Concept only, since there are better options.

Variations:

References:

References

1. NASA, **Brief History of Rockets**, web page 15 Apr 2012 (http://www.grc.nasa.gov/WWW/k-12/TRC/Rockets/history_of_rockets.html)
2. For a history of the rocket societies, see Frank Winter **Prelude to the Space Age**, Smithsonian Inst Press, 1983 (http://www.amazon.com/Prelude-Space-Age-Societies-1924-1940/dp/0874749638)

Section 2.4 - Thermal Engines

Strictly speaking combustion engines with expansion nozzles are also thermal engines since they rely on hot gases. For discussion purposes we placed that group on the previous page, and on this one look at non-combustion thermal engines. It is possible in theory to combine both into a single device, but this is not usually considered for practical design reasons. Performance is always important, and light molecules have higher exhaust velocity at a given temperature. Therefore the tendency is to use Hydrogen if practical since it is the lightest molecule. Liquid Hydrogen requires extremely cold temperatures (14K or -435F), so storage for long periods is not practical for a small tank. Larger systems have better surface to volume ratios and can use active cooling to keep it liquid.

Electro-thermal Engines

Electro-thermal methods convert externally supplied electric power to heating of the propellant.

36 Electric-Rail Rocket

Alternate Names: Electrothermal Ramjet

Type: Heated Gas Flow by Power Line

Description: High voltage electricity supplied by rails is shorted through a tungsten heat exchanger in the engine. This heats Hydrogen carried by the vehicle traveling between the rails. The rails are assumed to be set up on an incline. Performance would be the same as a nuclear thermal rocket, about 9 km/s exhaust velocity, since both use heated Hydrogen limited by the melting point of the engine components. It requires very high power levels for large vehicles: 44 MW / ton / g acceleration. Since power is coming from outside the vehicle, it can generate enough thrust for launch from Earth. Since the Hydrogen fuel is only stored for a short period, insulation and boiloff are not major problems. This method would compete with other ground accelerator type systems such as gas guns or electromagnetic coils.

Status: Concept only at this time.

Variations:

References:

- Wilbur, P. J.; Mitchell, C. E.; Shaw, B. D. "Electrothermal Ramjet", AIAA paper number 82-1216 presented at AIAA/SAE/ASME 18th Joint Propulsion Conference, Cleveland, OH, 21-23 June 1982.

37 Resistojet

Alternate Names:

Type: Heated Gas Flow by Photovoltaic Array

Description: In this method sunlight generates electricity, which is used to heat gas passed over or through a heating element, often after catalytic decomposition of a storable fuel to a lighter gas. In principle it is similar to the Electric rail rocket, but is used for smaller thrusters in orbit with attached solar array as power source. This limits the thrust, so it is not powerful enough to use for launch. Compared to chemical thrusters it gets about 50% better exhaust velocity.[1]

Status: This method was used to extend the operating life of communications satellites, since they had large solar arrays for their main job. The recent tendency is to use ion thrusters, which have even better performance.

Variations:

References:

- Louviere, Allen J. et al "Water-Propellant Resistojets for Man- Tended Platforms", NASA Technical Memorandum 100110, 1987.

Photo-thermal Engines

This group uses direct electromagnetic radiation (photons) from natural or artificial sources to heat the propellant.

38 Solar-Thermal Engine

Alternate Names:

Type: Heated Gas Flow by Solar Flux

Description: In this method sunlight is concentrated by a reflector or lens, and then heats an absorber. The absorber transfers heat to a working fluid, usually Hydrogen. The Hydrogen is then expanded through a nozzle. If the absorber is in the form of pipes, the exhaust velocity is limited to about 9 km/s. If the absorber is in the form of a particle bed, which does not require mechanical strength, refractory carbides can be used. Tantalum hafnium carbide, formula Ta_4HfC_5, is hypothesized to have a melting point of 4488 K (7619 F) and would set an upper limit to the particle bed method. There are obvious difficulties in testing that material as no other container could hold it when melted. The bed is rotated to keep the particles from being blown out, and Hydrogen flow is from the outside in, then out a nozzle. Sunlight is focused on the inside surface which is then the hottest point. Hydrogen dissociates above 3000K to single atoms, leading to exhaust velocities slightly above 10 km/s.

Solar concentrators can be very low mass, and use all of the solar spectrum. So they can reach higher power levels than electro-thermal thrusters. They are more suited to main propulsion rather than orbit maintenance as electro-thermal typically is. The direction of the Sun is usually not the same as the direction of thrust, and changes over time. So solar-thermal systems need a way to point the concentrators. One way to handle this is to roll the vehicle about the thrust axis plus pivot the concentrators about a perpendicular axis. The concentrators are typically large, so need to be assembled or unfolded in orbit.

Status: Components have been tested by the US Air Force. Ion and plasma thrusters have become preferred because of their 3-5 times higher exhaust velocity

Variations:

References:

- Gartrell, C. F. "Future Solar Orbital Transfer Vehicle Concept", IEEE Transactions on Aerospace Electronic Systems, vol AES-19 no 5 pp 704-10, 1983.

39 Laser-Thermal Engine

Alternate Names:

Type: Heated Gas Flow by Laser

Description: A laser beam from an external source is passed through a window into a chamber. It is then absorbed by a heat exchanger or is focused to create a laser-sustained plasma within the gas flow. Hot gas is then expelled through a nozzle. By using a powerful energy source external to the propellant, exhaust velocity of around 10 km/s can be reached with high thrust to mass. One method of doing this on Earth is with large, ground-based lasers. Alternately the lasers can be at ground level, and a directing mirror is located on top of a large tower. A vacuum pipe connects the two. The extra height avoids atmospheric distortion and allows more distance to the horizon. Use of laser propulsion only in an upper stage would allow smaller lasers than are required for a first stage system. Even so, powerful enough lasers are not available yet, which limits the use of the method. Other methods of supplying energy to a vehicle are likely to be less expensive.

Status: Concept only at present.

Variations:

References:

- Abe, T.; Shimada, T. "Laser Assisted Propulsion System Experiment on Space Flyer Unit", 38th International Astronautical Federation Conference paper number IAF-87-298, 1987.
- Abe, T.; Kuriki, K. "Laser Propulsion Test Onboard Space Station", Space Solar Power Review vol 5 no 2 pp 121-5, 1985.
- Jones, L. W.; Keefer, D. R. "NASA's Laser Propulsion Project", Astronautics and Aeronautics, v 20 no 9 pp 66-73, 1982.

40 Laser Detonation-Wave Engine

Alternate Names:

Type: Plasma via Laser

Description: In this method the propellant is a solid block with a flat bottom. A first laser pulse evaporates a layer of propellant. A second, larger, pulse creates a plasma detonation wave, which shocks and heats the propellant layer. The layer expands against the base of the solid block of remaining propellant. The pulse pattern is repeated as soon as the plasma dissipates and the laser can reach the block again. Because no fuel tank overhead or engine is needed, the vehicles are potentially very cheap, but it requires a powerful laser to function. For example, a 10 kg vehicle accelerated at 2 g's using a 20 km/s exhaust velocity plasma wave would require a pulsed laser with an average power of 2 MW, while large industrial pulsed lasers are about 600 W [2]. Since the propellant temperature is not limited by any container, it can be hotter than other thermal methods, and so have better performance.

10 kg is a feasible size for this type of vehicle, if the laser can be kept focused on it till it reaches orbit velocity. In theory this would deliver about 6 kg to orbit. For larger payloads, such as to carry passengers, the laser would need to scale to GigaWatt power levels, which has led to the common saying within space propulsion circles of "there is nothing wrong with laser propulsion except the lack of GigaWatt lasers". Similar power levels are generally required of all launch methods carrying metric ton or larger payloads from Earth. For example, the Space Shuttle Main Engines, of which the Space Shuttle Orbiter used three, each had a power of 9.2 GW

Status: Some detonation experiments have been carried out at small scale in a laboratory

Variations:

References:

- Kare, J.T. "SDIO/DARPA Workshop on Laser Propulsion, Volume 1: Executive Summary" Lawrence Livermore National Laboratory report number DE87-003254, 1987.

41 Microwave Thermal Engine

Alternate Names:

Type: Heated Gas Flow via Microwaves

Description: For this method microwaves from an external source are absorbed by a heat exchanger or concentrated by a waveguide into the engine. Hydrogen flows through the engine, absorbs the energy, and then exits by a nozzle. A large phased microwave array on the ground can focus onto a rocket-sized area over a range of hundreds of kilometers. Given a way to couple the microwave energy to a working fluid such as Hydrogen, this type of propulsion could provide significant launch vehicle velocities. High power microwave amplifiers exist in a variety of forms with efficiencies up to 75% and power levels up to one megawatt. As compared to Laser Thermal the chief advantage is the availability of high power microwave sources at relatively low cost. A disadvantage is the much larger wavelength of microwaves vs lasers, so maintaining focus over a distance is harder.

- Design Example: 10 meter diameter receiver, 5 cm wavelength, 1 km phased array, range = 200 km.

Status: Concept only so far.

Variations:

References:

Nuclear Thermal Engines

This group use a nuclear reactor to heat the propellant. They vary in the physical state of the reactor core (solid, liquid, or gas).

42 Solid Core Nuclear Engine

Alternate Names:

Type: Heated Gas Flow by Fission Reactor

Description: Hydrogen is heated by flowing through a critical **Nuclear Reactor**, then expelled through a nozzle at high velocity. The low molecular weight of Hydrogen gas allows a higher exhaust velocity, about 9 km/s, than combustion rockets. Advantages include high power levels and high total stored energy.

Issues for Nuclear Thermal include:

- For single missions the energy of the reactor has barely been tapped by the time the hydrogen is consumed, leaving a now radioactive core to work around if refueling for another mission.
- Radiation shielding is needed for crew and cargo. To some degree that is mitigated by the need for shielding from the natural space radiation environment.
- Any type of nuclear device raises extensive safety and environmental issues, even if not activated until in orbit, and even if not really warranted for technical reasons. One way around this issue is to mine and use the fissionable materials away from Earth. For example, parts of the Lunar surface have Thorium concentrations of 10 parts per million. If used as reactor fuel, that can provide a net energy of 350 MJ/kg of unprocessed Lunar soil, or about 7 times the energy density of gasoline on Earth. The processed fuel of course will be 100,000 times higher energy density, but the unprocessed ore energy density is an indicator of the feasibility of mining it.

In comparison to ion and plasma thrusters, nuclear thermal has about 3-5 times lower exhaust velocity, but much higher thrust levels. The near-instant burns relative to orbit time for nuclear thermal vs. constant burn for electric thrusters reduces the latter's advantage by 30%, but are still are 2.1-3.5 higher. The choice of method would depend on the importance of fuel mass, which is usually high, so today electric thrusters are usually preferred. In comparison to solar thermal it has about the same performance in terms of exhaust velocity, but higher thrust levels.

History The two major American reactor development efforts in the 1960s were KIWI and NERVA. Together with the $328 million spent on technology development, $90 million spent on the Nuclear Rocket Development Station in Nevada, and $153 million on other test facilities, almost $1.4 billion (in then-year dollars) was spent on nuclear rocket development from 1955 to 1972 (see **Appendix 2:Reference Data**). Although considerable engine testing was done, the problem of solid core fuel damage at high operating temperatures, which are desirable for performance, was not solved.

Status: Nuclear rockets reached the testing stage in the 1960's under the NERVA program. Lack of actual need in a mission and rising worries about anything nuclear led to a halt in development. Since then only minor studies have been done.

Variations:

- **LOX-Augmented Nuclear Thermal Rocket Propulsion** This injects Oxygen after the Hydrogen is heated by the reactor core [3]. This increases thrust by about a factor of 3, which is useful for initial launch, then transitions to pure Hydrogen later for higher efficiency. Exhaust velocity is lowered by about 1/3 when adding Oxygen due to the higher molecular weight of the resulting exhaust.
- **Particle Bed Nuclear Engine** - Although the nuclear rocket program was stopped a number of years ago, more recent work at Brookhaven National Laboratories on fluidized particle bed reactors warrants their consideration for launch vehicles. The small particle size (.3 mm) allows high heat transfer rates to the working fluid, hydrogen, and hence potentially high thrust to weight ratios. The smaller particles also potentially solves the fuel damage problem, as there is less scope for cracking in a fine powder. Exhaust velocity is increased slightly to 10 km/s.

References:

- Thomas, Ulrich "Nuclear Ferry - Cislunar Space Transportation Option of the Future", Space Technology (Oxford) v 7 no 3 pp 227-234, 1987.
- Holman, R.R.; Pierce, B. L. "Development of NERVA reactor for Space Nuclear Propulsion", presented at AIAA/ASME/SAE/ASEE 22nd Joint Propulsion Conference, Huntsville, Alabama, 16-18 Jun 1986, AIAA paper number 86-1582, 1986.
- Thom, K. et al "Physics and Potentials of Fissioning Plasmas for Space Power and Propulsion", Acta Astronautica vol 3 no 7-8 pp 505-16, Jul. -Aug. 1976.
- DiStefano, E. "Space Nuclear Propulsion - Future Applications and Technology", 2nd Symposium on Space Nuclear Power Systems, Albequerque, New Mexico, 14 January 1985, pp 331-342, 1987.

43 Liquid Core Nuclear

Alternate Names:

Type: Heated Gas Flow by Fission Reactor

Description: In order to attain higher performance than a solid core rocket, the reactor core is raised to a high enough temperature to become liquid. Hydrogen is bubbled through the liquid, then exhausted out a nozzle. The Hydrogen is first used to cool the reactor container, and so the temperature limit is governed by that rather than the melting point of the core. Expected exhaust velocity is up to 13-15 km/s, but development and testing of this type of engine will be difficult, as test failures can easily squirt the core fluid out the nozzle.

Status: Currently a concept only

Variations:

References:

44 Gas Core Nuclear

Alternate Names:

Type: Heated Gas Flow by Fission Reactor

Description: In this version the reactor core is hot enough to be in gaseous form. The Hydrogen flow is seeded with an absorbent material to directly absorb the thermal radiation from the core. The core is kept from leaking out the nozzle either by a transparent container (nuclear light bulb), a flow vortex, which uses the density difference between uranium and hydrogen, or by magnetic separation, which uses the ionization difference between the uranium and the hydrogen. Expected performance ranges from 15-20 km/s for a quartz container, up to 30-50 km/s for a flow vortex. The latter is in the range of ion or plasma thrusters, but development and testing would be as difficult as for liquid core due to the possibility of ejecting the core during tests. A solid core reactor producing electricity for an electric thruster would have the same performance as gas core nuclear, with fewer development problems.

Status: Currently a concept only.

Variations:

References:

- Wikipedia article: **Nuclear Lighbulb**

References:

1. Seitzman, J. **Electrothermal Thrusters** (http://soliton.ae.gatech.edu/people/jseitzma/classes/ae6450/electrothermal_thrusters.pdf), 2006.
2. JK Lasers Pulsed Nd:YAG Lasers spec sheet (http://www.jklasers.com/pulsed-nd-yag-lasers)
3. Next Big Future Blog **Recent Nuclear Thermal Rocket Proposals** (http://nextbigfuture.com/2009/06/nuclear-dc-x-recent-nuclear-thermal.html) 17 Jun 2009.

Section 2.5 - Bulk Matter Engines

This group of propulsion methods expel solid particles as reaction mass, or accelerate cargo directly.

45 Rotary Engine

Alternate Names:

Type:

Description: A one or two stage rotary mechanism mechanically accelerates a small amount of reaction mass, then releases it. In the two stage version, top speeds of ~6 km/s are possible. The advantage is to be able to use any available material as reaction mass. The disadvantage is the relatively low ejection velocity. There is an opportunity to mechanically augment another type of thruster by having it moving backwards relative to the vehicle center when running, thus adding up to 6 km/s to whatever the native exhaust velocity is. The duty cycle of the augmented thruster would be limited to about 25%, because the rest of the time it would be moving in the wrong direction to add velocity. This raises the issue of many on-off cycles of the thruster.

Status: Centrifuges and flywheels have long been in use. Versions optimized for ejecting mass at high velocity are still in the concept stage.

Variations:

References:

46 Coilgun Engine/Launcher

Alternate Names: Mass Driver Reaction Engine

Type:

Description: This is a type of electromagnetic accelerator. A carrier, or bucket, is accelerated by interaction of magnetic fields from 'driver' coils. The carrier holds a reaction mass/cargo, which is released. The bucket is slowed down and reused. There is no theoretical limit to velocity of this type of device - particle accelerators have reached near lightspeed - but practical limits are likely t be on the order of 10 km/s. As a vehicle engine it is not dramatically different than other electric thrusters in performance, aside from the benefit of using any bulk mass as propellant. As a fixed launcher for bulk materials, a large power supply and rapid recycling of the buckets allows very high mass delivery rates to orbit velocity. Unassisted bulk material will have an orbit that intersects the launch point, so a device in orbit will have to catch the cargo lots and add some velocity to them.

Status: This method was extensively studied as part of the "Space Colony" studies of the ~1970's, and prototype parts were built and tested.

Variations:

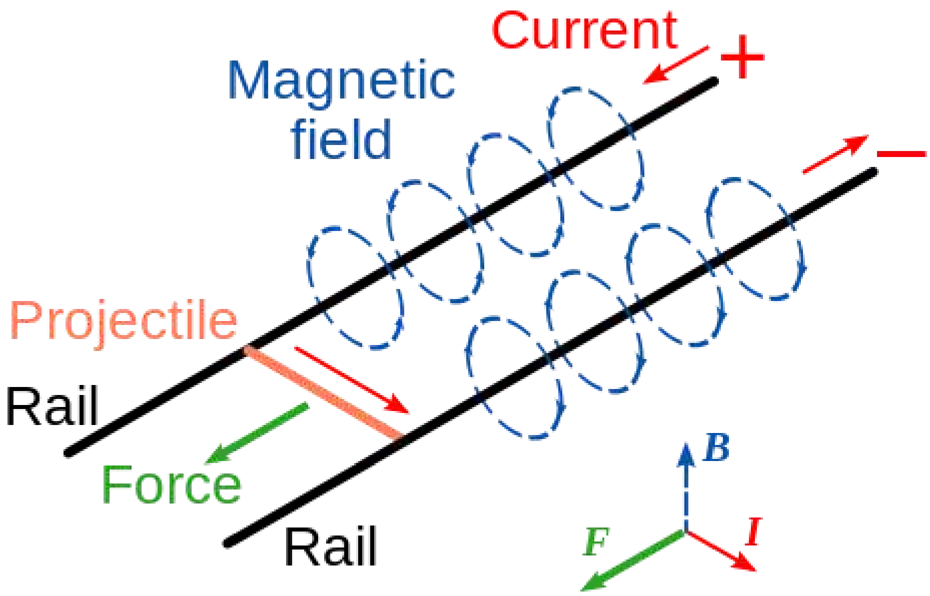
References:

47 Railgun Engine/Launcher

Alternate Names:

Type:

Description: A railgun consists of two current-carrying rails, and a projectile that creates a short circuit between the rails. The rails create a vertical magnetic field between them, as the current goes in opposite directions in each rail. The interaction of the field with the current in the short circuit creates a Lorentz force forward that accelerates a projectile. In a large version this would be used to launch vehicles in their entirety. In a smaller version it would eject a reaction mass and function as a propulsion engine.



Schematic of railgun operation.

Status: Railguns have been under development as replacements for conventional artillery, but are not yet operational as such. For launch vehicle acceleration, a much larger power supply would be needed, and this has so far been merely a concept

Variations:

Section 2.6 - Ion and Plasma Engines

Combustion type engines are limited in temperature, and thus gas exhaust velocity, either by the chemical energy of the fuel, or the melting point of the engine materials. Ion and plasma type engines bypass these limits and reach much higher velocities by one or more of: using external energy sources, lower energy density to limit engine heating, or using non-material containment such as magnetic fields to direct the flow. These methods also tend to make ion and plasma engines heavy compared to combustion engines, so they are not used for launch from high gravity bodies. Rather they operate where there is already a low-g environment in orbit or away from gravity wells.

The ion and plasma types both involve high energy particles. The distinction is one of density. In the former the particles act individually, while in the latter they are dense enough to require treating them as a flowing medium.

Ion Engines

48 Electrostatic Ion Engine

Alternate Names:

Type:

Description: Electrostatic ion thrusters generally work by first ionizing the propellant, then using a voltage gradient to accelerate the ions. They are capable of high exhaust velocities (30-50 km/s), but relatively low energy density and thrust in order to prevent overheating and erosion of the screens creating the voltage difference. Since the voltage gradient is uniform between two flat screens, the ion beam is well directed without a nozzle. Ejecting only positive ions would produce a net charge on the engine, so ion engines also have an electron gun to balance the charges emitted.

All ion engines require an external power source. A number of variations for the power source have been proposed, but only Solar-Electric has actually been used. Since ionization merely removes an electron, but does not accelerate the ion, it represents an efficiency loss for the engine. Therefore ion engines tend to prefer fuels with high atomic weight relative to ionization energy. Since most high atomic weight elements are solids, Mercury and Xenon have often been chosen, and sometimes Argon if operating in a cold environment or reaching a very high exhaust velocity is desired.

Status: Ion thrusters have been used in recent years on some communications satellites and interplanetary spacecraft.

Variations:

- **48a Solar-Electric Ion** - Sunlight is converted to electricity by a photovoltaic array. The electricity is used to ionize and accelerate the propellant by electrostatic voltage.
- **48b Thermoelectric Ion** - Radioactive isotope decay produces heat. Heat is converted to electricity by semiconductors. Electricity ionizes and accelerates atoms in the engine.
- **48c Laser-Electric Ion** - A laser tuned to the optimum absorption wavelength of photovoltaic cells supplies power. The cells convert laser light to electricity, which is used to power the ion engine. The ion engine then ionizes and accelerates the propellant. The reason for using a laser is to get higher power levels than natural sunlight provides.
- **48d Microwave-Electric Ion** - A microwave receiving antenna, or **Rectenna**, on the spacecraft converts microwaves to electricity. Electricity is used to ionize and accelerate atoms. Microwave antennas can be low mass, but the operating distance is limited by the ability to focus the long wavelength.

- **48e Nuclear-Electric Ion** - A nuclear reactor generates heat, which is converted to electricity in thermoelectric or turbine/generator cycles. Electricity is used to ionize propellant and accelerate it by electrostatic voltage. This is more suited to high power applications, and outer Solar System locations where sunlight is weak.
- **48f Electrospray Thruster** - An ionic liquid is emitted from a capillary nozzle similar to an inkjet printer. A high voltage electrode then accelerates the charged droplets. This is suited to very small thrust levels for small spacecraft, or for steering, since lithography can be used to create microscopic nozzles.
- **48g Colloid Thruster** - Rather than individual ions, this method accelerates relatively massive colloidal charged particles.

References:

Solar-Electric, 1970s

- Loeb, H. W. "Electric Propulsion Technology Status and Development Plans - European Programs (Space Vehicles)", J. Spacecraft and Rockets , vol 11 no 12 pp 821-8, Dec. 1974.
- Mutin, J.; Tatry, B. "Electric Propulsion in the Field of Space", Acta Electron. (France) vol 17 no 4 pp 357-70, Oct. 1974 (in French).
- Byers, D. C.; Rawlin, V. K. "Critical Elements of Electron- Bombardment Propulsion for Large Space Systems", J. Spacecraft and Rockets vol 14 no 11 pp 648-54, Nov. 1977.
- Parkash, D. M. "Electric Propulsion for Space Missions", Electr. India vol 19 no 7 pp 5-15, 15 April 1979.

Solar-Electric, 1980s

- Kaufman, H. R. "Performance of Large Inert-Gas Thrusters", AIAA paper number 81-0720 presented at 15th International Electric Propulsion Conference, Las Vegas, Nevada, 21-23 April 1981.
- Clark, K. E.; Kaufman, H. B. "Aerospace Highlights 1981: Electric Propulsion", Astronautics and Aeronautics, v 19 no 12 pp 58-59, 1981.
- Zafran, S. et al "Aerospace Highlights 1982: Electric Propulsion", Astronautics and Aeronautics, v 20 no 12 pp 71-72 1982.
- James, E.; Ramsey, W., Sr.; Steiner, G. "Developing a Scaleable Inert Gas Ion Thruster", AIAA paper number 82-1275 presented at AIAA/SAE/ASME 18th Joint Propulsion Conference, Cleveland, OH, 21- 23 June 1982.
- Anon. "Ion Propulsion Engine Tests Scheduled", Aviation Week and Space Technology, v 116 no 26 pp 144-5, 1982.
- Brophy, J. R.; Wilbur, P. J. "Recent Developments in Ion Sources for Space Propulsion", Proceedings of the Intl. Ion Engineering Congress vol 1 pp 411-22, 1983.
- Bartoli, C. et al "Recent Developments in High Current Liquid Metal Ion Sources for Space Propulsion", Vacuum vol 34 no 1-2 pp 43-6, Jan. -Feb. 1984.
- Jones, R. M.; Poeschel, R. L. "Primary Space Propulsion for 1995-2000 - Electrostatic Technology Applications" AIAA/SAE/ASME 20th Joint Propulsion Conference, AIAA paper number 84-1450, 1984.
- Imai, R.; Kitamura, S. "Space Operation of Engineering Test Satellite -III Ion Engine", Proceedings of JSASS/AIAA/DGLR 17th Intl. Electric Propulsion Conf. pp 103-8, 1984.
- Bartoli, C. et al "A Liquid Caesium Field Ion Source for Space Propulsion", J. Phys. D vol 17 no 12 pp 2473-83, 14 Dec. 1984.
- Voulelikas, G. D. "Electric Propulsion: A Review of Future Space Propulsion Technology" Communications Research Centre, Ottawa, Ontario, report number CRC-396, October 1985.
- Nakamura, Y.; Kuricki, K. "Electric Propulsion Test Onboard the Space Station", Space Solar Power Review vol 5 no 2 pp 213-9, 1985.
- Rawlin, Vincent K; Patterson, Michael J. "High Power Ion Thruster Performance", NASA Technical Memorandum 100127, 1987.
- Mitterauer, J. "Liquid Metal Ion Sources as Thrusters for Electric Space Propulsion", J. Phys. Colloq. (France) vol 48, no C-6, pp 171-6, Nov. 1987.
- Mitterauer, J. "Field Emission Electric Propulsion - Emission Site Distribution of Slit Emitters", IEEE Trans. on Plasma Sci. vol PS-15, pp 593-8, Oct. 1987.
- Stuhlinger, E. et al "Solar-Electric Propulsion for a Comet Nucleus Sample Return Mission" presented at 38th Congress of the International Astronautical Federation, Brighton, England, 10 October 1987.

Laser-Electric Ion:

- Maeno, K. "Advanced Scheme of CO2 Laser for Space Propulsion", Space Solar Power Review vol 5 no 2 pp 207-11, 1985.

Microwave-Electic Ion:

- Nordley, G. D.; Brown, W. C. "Space Based Nuclear-Microwave Electric Propulsion", 3rd Symposium on Space Nuclear Power Systems, Albuquerque, New Mexico, 13 January 1986, pp 383-95, 1987.

Nuclear-Electric Ion:

- Reichel, R. H. "The Air-Scooping Nuclear-Electric Propulsion Concept for Advanced Orbital Space Transportation Missions", J. British Interplanetary Soc. vol 31 no 2 pp 62-6, Feb. 1978.
- Hsieh, T. M.; Phillips, W. M. "An Improved Thermionic Power Conversion System for Space Propulsion", Proceedings of the 13th Intersociety Energy Conversion Engineering Conference pp 1917-1923, 1978.
- Ray, P. K. "Solar Electric versus Nuclear Electric Propulsion in Geocentric Space", Trans. Am. Nucl. Soc. vol 39 pp 358-9, Nov.-Dec. 1981.
- Powell, J. R.; Botts, T. E.; Myrabo, L. N. "Annular Bed Nuclear Power Source for Electric Thrusters", AIAA paper number 82-1278 presented at AIAA/SAE/ ASME 18th Joint Propulsion Conference, Cleveland, Ohio, 21-23 June 1982.
- Powell, J. R.; Boots, T. E. "Integrated Nuclear Propulsion/Prime Power Systems", AIAA paper number 82-1215 presented at AIAA/SAE/ASME 18th Joint Propulsion Conference, Cleveland, Ohio, 21-23 June 1982.
- Buden, D.; Garrison, P. W. "Space Nuclear Power Systems and the Design of the Nuclear Electric Propulsion OTV", presented at AIAA/SAE/ASME 20th Joint Propulsion Conference, AIAA paper number 84-1447, 1984.
- Cutler, A. H. "Power Demands for Space Resource Utilization", Space Nuclear Power Systems 1986 pp 25-42.

Electrospray Thruster:

- **Micromachined propulsion systems for very small satellites**, EFPL Microsystems for Space Technologies Lab.

Plasma Engines

49 Arc Jet Engine

Alternate Names:

Type:

Description: Sunlight is converted to electricity by a photovoltaic array. The electricity is arced through a propellant stream, heating it to a plasma. The propellant is then expanded through a nozzle. No specific protection from heating is used in the nozzle. The arc jet method uses relatively low energy density and thrust levels, and a thick-walled refractory metal for the chamber and nozzle. Exhaust velocities of about 6-8 km/s are reached with liquid fuels like Hydrazine or Ammonia.

Status: Arc Jets have been used on spacecraft, and are commonly used on Earth as plasma torches for cutting metals.

Variations:

- **Pulsed Plasma Thruster** - operates by ablating Teflon material between two electrodes by means of an electric arc. The Lorentz force accelerates the material out as thrust.
- **Vacuum Arc Thruster** - operates by vaporizing and ionizing cathode material by means of a vacuum arc. The plasma accelerates outward as thrust.
- **Magnetoplasmadynamic Thruster** - operates via electric arc between central cathode and shell anode. Flowing gas between them is ionized and accelerated by the Lorentz force from current and field interactions.

References:

- Hardy, Terry L.; Curran, Francis M. "Low Power DC Arcjet Operation with Hydrogen/Nitrogen/Ammoinia Mixtures", NASA Technical Memorandum 89876, 1987.
- Stone, James R.; Huston, Edward S. "NASA/USAF Arcjet Research and Technology Program", NASA Technical Memorandum 100112, 1987.
- Kagaya, Y. et al "Quasi-steady MPD Arc-jet for Space Propulsion", Symposium for Space Technology and Science, Tokyo, Japan, 19 May 1986, pp 145-154, 1986.
- Manago, Masata et al "Fast Acting Valve for MPD Arcjet", IHI Engineering Review, v 19 no 2 pp 99-100, April 1986.

- Pivirotto, T. J.; King, D. Q. "Thermal Arcjet Technology for Space Propulsion", Chemical Propulsion Information Agency, Laurel, Maryland, 1985.

50 Electron Beam Heated Plasma Engine

Alternate Names:

Type:

Description: A high voltage (hundreds of keV) electron beam is injected axially into a propellant flow. The electron beam heats the flow to plasma temperatures, which produces high specific impulse. Cool gas is injected along the chamber walls to provide film cooling and protect the chamber from the very high temperature plasma.

Status:

Variations:

References:

51 Microwave Heated Plasma Engine

Alternate Names: Electron-Cyclotron Absorption Rocket, Variable Specific Impulse Magnetoplasma Rocket (VASIMR), Helicon plasma thruster

Type:

Description: Partially ionized gas directly absorbs microwaves, becoming hot, then expands through rocket nozzle. To keep the hot plasma from damaging the engine, a magnetic field is used for confinement, often using superconductors for efficiency. Current versions use Argon as propellant, but other gases should function with tuning. It may be possible to use unprocessed rock as fuel.

Status: VASIMR is in ground testing

Variations:

References:

- Personal communication from Ad Astra Rocket Company in reference to alternate fuels, 27 Mar 2012. Extensive publications about their thruster are linked to the main description page.

52 Fusion Heated Plasma Engine

Alternate Names:

Type:

Description: The exhaust of a pure fusion rocket is a thin, extremely hot plasma which gives very high performance. If higher thrust is needed, Hydrogen can be mixed with the plasma. This increases thrust at the expense of performance. By varying the mixture ratio, the performance vs thrust can be adjusted as needed during a mission.

Status: Fusion engines await practical fusion reactors, which are still in the research stage as of 2012.

Variations:

- **Reactor Leakage Mixed** - Some of the fusion reaction plasma leaks past the containment fields. This may be a mix of reacted particles, and un-reacted fuel. The leakage can be directly used as engine exhaust, or further mixed with additional material for higher mass flow/thrust at lower performance.
- **Plasma Kernel Mixed** - The fusion core plasma may be intentionally seeded with non-reacting material, which gets heated as part of the reaction. A certain percentage is directed out of the core for thrust, balanced

by new fuel added to the core.The possible advantage of this method is eliminating the mixing problem of adding mass after the reactor Fusion engines produce very high exhaust velocities, so trying to mix more matter into the stream may require very large engine components.Otherwise the stream may simply be gone before it has a chance to mix.

References:

53 Antimatter-Heated Plasma Engine

Alternate Names:

Type:

Description: The exhaust of a pure antimatter rocket is charged particles and gamma rays. This gives an extraordinarily high exhaust velocity, but relatively low thrust. If higher thrust is needed, hydrogen can be mixed with the plasma, at at the expense of performance. This likely requires a large magnetic bottle to contain the particles and Hydrogen ions (protons) long enough to mix.

Status:

Variations:

References:

53 Electrodeless Lorentz Force (ELF) Thruster

Alternate Names:

Type:

Description: The ELF-160A thruster creates a high-density, magnetized plasmoid known as a Field Reversed Configuration (FRC) employing a Rotating Magnetic Field (RMF). The RMF driven azimuthal currents, coupled with the enhanced axial magnetic field gradient produced by the FRC inside the flux preserving conical thruster, produce a large axial JθxBr force that accelerates the plasmoid to high velocity. The ELF-160A is completely electrodeless, the propellant is magnetically isolated from the thruster body, quasi-neutral, and there is zero contact between high temperature propellant and the thruster

Status: Active Development under Department of Defense contract.

Variations: One variation is the ElectroMagnetic Plasmoid Thruster (EMPT), an electrodeless pulsed plasma thruster that generates and accelerates a Field Reversed Configuration (FRC) to produce thrust.

References:

- Slough, J.; Kirtley, D.; Weber T. "Pulsed Plasmoid Propulsion: The ELF Thruster", presented at the 31st International Electric Propulsion Conference, September 20-24 2009
- Pancotti, A. P.; Little, J. M.; Neuhoff, J. S.; Cornella, B. M.; Kirtley D. E.; Slough J. T "Elecrodeless Lorentz Force (ELF) Thruster for ISRU and Sample Return Mission", presented at Joint Conference of 30th International Symposium on Space Technology and Science, 34th International Electric Propulsion Conference and 6th Nano-satellite Symposium, July 4-10 2015

Further Reading

- Wikipedia: plasma propulsion engine

Section 2.7 - High Energy Particle Engines

Particle engines generally have very good performance because of the high velocity of the particles. The high velocity also requires a lot of energy, since particle kinetic energy goes as the square of the velocity

A. Particle Rockets

Particle rockets are distinguished by emitting particles from an internal source, while the next section, External Particle Interactions, covers methods that employ particles in the environment, either natural or man-made.

54 Pulsed Fission Nuclear Thruster

Alternate Names: Orion

Type: Atomic Particles by Nuclear Detonation

Description: A series of small atomic bombs yield debris/particles which push against a plate/shock absorber arrangement. The shock absorber evens out the explosion pulses to a steady acceleration for the vehicle. Issues with this method are not so much technical feasibility as risks of using it near Earth and having a shipload of nuclear weapons, not to mention the radiation levels for crew and cargo. Benefits are the enormous potential payload and high acceleration. A possible application is moving dangerous asteroids. That would not take very many bombs, and the pusher plate/shock absorber could be made from asteroidal metals, even from the asteroid you are moving.

Other applications might involve moving large amounts of material in the outer Solar System, or starting interstellar trips. Far away from Earth the issue of radiation added to the environment is not as significant, since solar and cosmic radiation are already present i large amounts.

Status: Atomic bombs are well tested (unfortunately). A demonstration of the pusher plate technology was done in small scale with conventional explosives.

Variations:

> **54a Antimatter Catalyzed Pulsed Nuclear** - This speculative method uses antimatter triggered implosion of plutonium targets. In theory this would lower the critical mass to ~ 2 grams significantly lowering the size threshold for pulsed nuclear.
> **54b Implosion Driver Pulsed Nuclear** - Standard fission bombs are triggered by a chemical implosion. This method substitutes a laser, ion beam, or other external driver to collapse the fuel load. It potentially reduces the critical mass with a stronger collapse than chemicals can achieve, or by using a different isotope.

References:

55 Inertial Confinement Nuclear Engine

Alternate Names: Microfusion Engine

Type: Atomic particles by Fission or Fusion Reactor

Description: There are two major methods being pursued for fusion power. Magnetic confinement holds a hot plasma in a more or less steady state long enough for fusion reactions to produce net power. Inertial methods use lasers or particle beams to compress a fuel pellet to high pressure and temperature such that the fusion reaction happens very fast, after which it rapidly expands. Both methods can be applied to space propulsion. Magnetic confinement was addressed at **Fusion Heated Plasma Engine**, and this method addresses Inertial confinement.

A fuel pellet consists of a fusion core material, such as a Deuterium/Tritium mix, surrounded by a liner to optimally absorb the laser or particle beam energy. Optionally a fission shell surrounds the fusion core. This is similar to the arrangement of a fusion atomic bomb. Alternately particle beams can consist of fissionable heavy ions or fusible light element ions, or even highly accelerated solid pellets of inert or reactive fuels. The end result in all cases is a rapidly expanding cloud of highly energetic particles. These are directed by a shaped magnetic field or pusher plate to produce thrust. Thrust can be varied by how often you generate the explosions.

Status: Inertial confinement fusion is being researched for power generation. A chemical implosion is also the method by which most nuclear bombs are set of.

Variations: Multiple variations are possible by combining:

- Compression sources: laser, particle/ion beams, or solid pellets,
- Type of central target: none in the case of colliding pellets, fission and/or fusion fuel ingredients,
- Type of pellet liner: none or what material, and
- Thrust method: physical or magnetic nozzle

References:

56 Alpha Particle Emitter

Alternate Names:

Type: Atomic Particles by Radioactive Decay

Description: A radioactive element coats one side of a thin sheet which is capable of absorbing alpha particles. The particles emitted into the sheet are absorbed, while the particles emitted in the opposite direction escape, providing net thrust. Advantages are this is a simple device, and heating of the sheet can produce power as a side effect via thermoelectric or heat engine. Disadvantages are low thrust, and you cannot turn off radioactivity. You would have to close two plates so they face each other to neutralize thrust, and open them like a book to turn the thrust on.

Status: Alpha emitters are well known.Using them for propulsion is theoretical so far

Variations: Choice of radioactive material allows a wide range thrust levels.

References:

57 Fission Fragment Engine

Alternate Names:

Type: Atomic Particles by Fission Reactor

Description: Thin wires containing fissionable material are arranged to allow the nuclear fragments from the fission to escape. They are allowed to decay naturally for low thrust devices, or arranged in a nuclear critical arrangement for high thrust. The fragments are aimed by electrostatic or electromagnetic fields to mostly go out the back end of the thruster. The performance is very high because of the high speed of the fragments. The fission decay tends to damage whatever the wires are made of, so provision for replacing or reforming the wires would be needed for long term use.

Status: Currently theoretical only

Variations:

- **57a Antimatter Sail** - Antiprotons are directed at a sail coated with fissionable material such as Uranium. Their negative charge allows them to be captured in the electron shell of the atoms, and then causes fission by annihilation with one of the nuclear protons. One fission fragment is absorbed by the sail, while the other moves away at several percent of the speed of light, generating thrust.

References:

- Dusty Plasma Based Fission Fragment Nuclear Reactor Clark and Sheldon, Joint Propulsion Conference, 2005.

58 Pure Fusion Engine

Alternate Names:

Type: Atomic Particles by Fusion Reactor

Description: Various thermonuclear fusion reactors have been proposed. The results of a fusion reaction are high energy particles which can, in principle, be harnessed for propulsion. This differs from 52 Fusion Heated Plasma Engine in that no additional fuel is added beyond the fusion reactants. Instead the high temperature plasma or escaping charged particles are used directly as rocket exhaust. Since a break even fusion reaction has not been achieved as of yet, this remains a theoretical concept.

The actual method of containing the fusion plasma will influence the design of the rocket and its feasibility, in particular the mass of the device. A tokamak or stellarator plasma containment device would present the largest problems in siphoning off plasma for propulsion and their large weight would lead to an uninspiring thrust to mass ratio. Inertial confinement fusion using high power lasers or x-rays on small targets would probably be easier to build into a propulsion device, however the enormous size of the lasers and x-ray generators of current internal fusion projects would lead to the same problem with the thrust to mass ratio.

Status: Significant research is in progress for fusion in general. Applications to space propulsion are theoretical. Besides the large funding for magnetic and inertial confinement fusion, there are several alternate approaches that have lower funding levels.

Variations:

- **58a Magnetic Confinement** - Plasma in a chamber similar to a tokamak fusion power reactor is intentionally leaked to a magnetic nozzle.
- **58b Inertial Confinement** - The fuel pellet is heated and compressed by lasers, electron beam, or ion beam. After fusing, the resulting plasma is directed by a magnetic nozzle.
- **58c Electrostatic Confinement** - The fusion fuel is confined by a spherical potential well of order 100 kV. When the fuel reacts, the particles are ejected with energy of order 2 MeV, so escape the potential well. The potential well is at the focus of a paraboloidal shell, which reflects the fusion particles to the rear in a narrow beam (20-30 degree width).
- **58d Plasma Mantle Confinement** - The fusion fuel is contained in a toroidal/poloidal current pattern, similar to a Tokamak except all the currents are in the plasma. The current pattern is surrounded by a plasma sheath which isolates the fuel from a surrounding working fluid. The fluid provides mechanical compression, which heats the fuel to fusion ignition. After the fuel burn is completed, the energy generated heats the working fluid to high temperature, which then goes out a nozzle producing thrust.
- **58e Dense Plasma Focus** - A high current discharge in a radial arrangement causes the plasma to collapse to high temperature and pressure. This is being developed by **Lawrenceville Plasma Physics** in Middlesex,

New Jersey. This device has a reported fusion energy of 0.044 Joules vs capacitor energy of 50 kJoules. The output vs input energy ratio is a measure of how close the device is to practical operation.

- **58f Magnetized Target Fusion** - Colliding plasma balls are further compressed by acoustic shock waves generated by mechanical drivers. **General Fusion** near Vancouver, Canada is researching this method.

References:

- Freeman, M. "Two Days to Mars with Fusion Propulsion", 21st Century Science and Technology, vol 1, pp 26-31, Mar.-Apr. 1988.
- Kammash, T.; Galbraith, D. L. "A Fusion-Driven Rocket Propulsion Scheme for Space Exploration", Trans. Am. Nucl. Soc. vol 54 pp 118-9, 1987.
- Mitchell, H. M.; Cooper, R. F.; Verga, R. L. "Controlled Fusion for Space Propulsion. Report for April 1961-June 1962", US Air Force report number AD-408118/8/XAB, April, 1963.

Inertial Confinement

- Kammash, T.; Galbraith, D. L. "A Fusion Reactor for Space Applications", Fusion Technology, v. 12 no. 1 pp 11-21, July 1987.
- Orth, C. D. et al "Interplanetary Propulsion using Inertial Fusion", report number UCRL--95275-Rev.1: 4th Symposium on Space Nuclear Power Systems, Albequerque, New Mexico, 12 January 1987.
- Hyde, Roderick, "A Laser Fusion Rocket for Interplanetary Propulsion" , LLNL report UCRL-88857. Topics include:

  - Fusion Pellet design: Fuel selection. Energy loss mechanisms. Pellet compression metrics.
  - Thrust Chamber: Magnetic nozzle. Shielding. Tritium breeding. Thermal modeling. Fusion Driver (lasers, particle beams, etc): Heat rejection.
  - Vehicle Summary: Mass estimates.
  - Vehicle Performance: Interstellar travel required exhaust velocities at the limit of fusion's capability. Interplanetary missions are limited by power/weight ratio. Trajectory modeling. Typical mission profiles. References, including the 1978 report in JBIS, "Project Daedalus", and several on ICF and driver technology.

- Bussard, Robert W., "Fusion as Electric Propulsion", Journal of Propulsion and Power, Vol. 6, No. 5, Sept.-Oct. 1990. ***Abstract:*** Fusion rocket engines are analyzed as electric propulsion systems, with propulsion thrust- power-input-power ratio (the thrust-power "gain" G(t)) much greater than unity. Gain values of conventional (solar, fission) electric propulsion systems are always quite small (e.g., G(t)<0.8). With these, "high-thrust" interplanetary flight is not possible, because system acceleration (a(t)) capabilities are always less than the local gravitational acceleration. In contrast, gain values 50-100 times higher are found for some fusion concepts, which offer "high-thrust" flight capability. One performance example shows a 53.3 day (34.4 powered; 18.9 coast), one-way transit time with 19% payload for a single-stage Earth/Mars vehicle. Another shows the potential for high acceleration (a(t)=0.55g(o)) flight in Earth/moon space.)

Electrostatic Confinement

- Bussard, Robert W., "The QED Engine System: Direct Electric Fusion-Powered Systems for Aerospace Flight Propulsion" by Robert W. Bussard, EMC2-1190-03, available from Energy/Matter Conversion Corp., 9100 A. Center Street, Manassas, VA 22110. Summary: This is an introduction to the application of Bussard's version of the Farnsworth/Hirsch electrostatic confinement fusion technology to propulsion. 1500<Isp<5000 sec. Farnsworth/Hirsch demonstrated a 10**10 neutron flux with their device back in 1969.

- **Plasma Mantle Confinement**

- Koloc, Paul M., "PLASMAKtm Star Power for Energy Intensive Space Applications", Eighth ANS Topical Meeting on Technology of Fusion Energy Fusion Technology , March 1989.

This note is 20 years old and needs updating: Aneutronic energy (fusion with little or negligible neutron flux) requires plasma pressures and stable confinement times larger than can be delivered by current approaches. If plasma pressures appropriate to burn times on the order of milliseconds could be achieved in aneutronic fuels, then high power densities and very compact, relatively clean burning engines for space and other special applications would be at hand. The PLASMAK[a] innovation will make this possible; its unique pressure efficient structure, exceptional stability, fluid-mechanically compressible Mantle and direct inductive MHD electric power conversion advantages are described. Peak burn densities of tens of megawats per cc give it compactness even in the multi-gigawatt electric output size. Engineering advantages indicate a rapid development schedule at very modest cost. (I strongly recommend that people take this guy seriously. Bob Hirsch, the primary proponent of the Tokamak, has recently declared Koloc's PLASMAKª precursor, the spheromak, to be one of 3 promising fusion technologies that should be pursued rather than Tokamak. Aside from the preceeding appeal to authority, the PLASMAKª looks like it finally models ball-lightning with solid MHD physics. -- Jim Bowery)

  - **Dense Plasma Focus**

- Lerner, Eric et al, **Fusion Reactions from >150 keV Ions**, *Physics of Plasmas*, v 19 issue 3, March 2012.

59 Neutral Particle Beam Thruster

Alternate Names:

Type:

Description: A high energy (order 50 MeV) particle accelerator generates a proton beam. This beam is neutralized (combined with electrons to make neutral atoms), then ejected. The exhaust is moving at a substantial fraction of the speed of light, so performance is very high. This type of machine was explored under the US missile defense program as a way of destroying missiles (with the beam). The energy required for space propulsion with this method exceeds the energy available from nuclear fusion, so it only makes sense with antimatter or external power sources such as a laser

Status: Particle accelerators of this energy have existed since the mid-20th century. What limits this method is lack of a power supply, so it remains untested.

Variations:

- **59a Near Lightspeed Probe** - If the particle accelerator produces very high energy protons, such that relativistic mass increase is significant, and the power source is a very powerful laser located at the origin star, and focused using the star as a gravitational lens, the rocket equation no longer constrains the final velocity. Velocities close to the speed of light would be possible, and stopping would be possible by pointing the accelerator in the opposite direction. Acceleration in this case is limited by the power to mass ratio of the power conversion and particle accelerator hardware.

References:

60 Antimatter Annihilation

Alternate Names:

Type:

Description: Atoms and anti-atoms (or their constituent particles) annihilate, producing pions, then muons, then gamma rays. The charged particles can be acted upon by a magnetic nozzle. The gamma rays can be absorbed by a container, and the resulting heat used to supply power. Antimatter provides the highest theoretical energy fuel (100% matter to energy conversion), although the overhead involved with storing antimatter may reduce the practical efficiency to a level comparable to other propulsion methods.

Status: Small amounts of antimatter have been created and temporarily stored at particle accelerator labs. At present, antimatter is not thought to exist naturally in large quantities.

Variations:

- Wikipedia article: **Redshift Rocket**

References:

- Forward, Dr. Robert L. **Antiproton Annihilation Propulsion**, AFRPL TR-85-034 from the Air Force Rocket Propulsion Laboratory (AFRPL/XRX, Stop 24, Edwards Air Force Base, CA 93523-5000). NTIS AD-A160 734/0 **Note:** This is a technical study on making, holding, and using antimatter for near-term (30-50 years) propulsion systems. Excellent bibliography. Forward is the best-known proponent of antimatter. This also may be available as

UDR-TR-85-55 from the contractor, the University of Dayton Research Institute, and DTIC AD-A160 from the Defense Technical Information Center, Defense Logistics Agency, Cameron Station, Alexandria, VA 22304-6145. And it's also available from the NTIS, with yet another number

- G. D. Nordley, **Application of Antimatter - Electric Power to Interstellar Propulsion**, Journal of the British Interplanetary Society, June 1990.

B. External Particle Interactions

61 Magsail

Alternate Names:

Type:

Description: The magsail operates by placing a large superconducting loop in the solar wind or planetary ion stream. The current loop produces a magnetic field that deflects the ions, producing a reaction force. Because of the large area covered by the loop, which is mostly empty, it can develop a relatively large force for a given mass. It is limited in direction and strength by the local solar wind or ion flow.

Status: Magnetic field interaction with an ion stream is well understood. Practicality of this method depends on an efficient enough design. Has not been tested as of 2012.

Variations:

References:

62 External Particle Beam

Alternate Names:

Type:

Description: A remote particle beam source aims it at a target vehicle. The particles are either absorbed or reflected to generate thrust directly, or their kinetic energy is used as a power source for some other type of propulsion. The major issue is keeping the beam narrow enough to be a useful energy transfer method.

Status: Untested as of 2012

Variations:

References:

63 Interstellar Ramjet

Alternate Names: Bussard Ramjet

Type:

Description: A large funnel or inlet is used to compress interstellar gas, which is then fed to a fusion reactor for propulsion. Because of the low density of the interstellar medium, an extraordinarily large scoop is required to get any useful thrust. Performance is limited by the exhaust velocity of the fusion reaction to a few percent of the speed of light. In other words, collecting the gas causes drag, and at the exhaust velocity of the reactor there will be no net thrust. This makes the system an efficient decelerator by pointing the exhaust forward. In that case the drag and reverse thrust both act to slow the vehicle. Variable density of interstellar gas affects the viability of this method. By running the reactor partly off of internal fuel, this type of vehicle can be brought to high enough velocity for the collector to start to function, and then to go to higher velocities. By having the reactor use less fuel than collected, it can be self-refueling.

Status: Although the concept dates back to 1960, it is untested due to a lack of practical fusion reactors.

Variations:

References:

[D77] R. W. Bussard, "Galactic Matter and Interstellar Flight", Astronautica Acta 6 (1960): 179 - 194.

[D78] A. R. Martin, "The Effects of Drag on Relativistic Spacefight", JBIS 25 (1972):643-652

[D79] N. H. Langston, "The Erosion of Interstellar Drag Screens", JBIS 26 (1973): 481-484.

[D80] D.P. Whitmire, "Relativistic Spaceflight and the Catalytic Nuclear Ramjet", Acta Astronautica 2 (1975): 497 - 509.

[D80] C. Powell, "Flight Dynamics of the Ram-Augmented Interstellar Rocket", JBIS 28 (1975):553-562

[D81] D.P. Whitmire and A.A. Jackson, "Laser Powered Interstellar Ramjet", JBIS 30 (1977):223 - 226.

[D82] G. L. Matloff and A. J. Fennelly, "Interstellar Applications and Limitations of Several Electrostatic/Electromagnetic Ion Collection Techniques", JBIS 30 (1977):213-222

64 Interstellar Scramjet

Alternate Names:

Type:

Description: Similar to the interstellar ramjet, the interstellar medium is compressed to fusion density and temperature. In this concept it is only compressed laterally, then re-expanded against a nozzle. Incredible vehicle sizes and lengths are required to reach fusion conditions, but since lateral compression does not cause as much drag as full compression, it is not limited by the exhaust velocity. Thus speed may reach a substantial fraction of the speed of light.

Status:

Variations:

References:

65 External Fuel Supply

Alternate Names:

Type:

Description: Any propulsion system which stores all it's fuel at the start has an exponential fuel requirement as a function of velocity. If added fuel is delivered by a particle or pellet stream, either from behind or positioned ahead, that exponential requirement is turned into a linear one. This is because at each point the vehicle is only accelerating itself (plus a small amount of fuel), instead of the vehicle plus all the fuel for the entire mission. In the latter case you are using fuel early to accelerate fuel for later, which causes the exponential overhead.

Status:

Variations: Lots of variations are possible in terms of what fuel is delivered, and if it is formed into a particle stream, discrete pellets or container tanks of fuel. Incidentally this can also be used for other types of supplies than fuel, but that belongs in the Engineering Methods section.

References:

Further reading

- "So You Wanna Build A Rocket?" at Project Rho gives quick estimates for a variety of atomic rockets -- including the Bussard Ramjet, the Nuclear Lightbulb, the Nuclear Salt Water Rocket, the magsail, etc. -- and related engineering challenges -- including radiation sheilding, space suits, etc.
- "The Relativistic Rocket" in the Physics FAQ describes some of the relativistic effects of traveling close to the speed of light, and one way to calculate "How much fuel is needed?".

Section 2.8 - Photon Engines

There are a variety of propulsion techniques that use either internally generated or externally supplied photons for propulsion. Stars, of course, are a natural supply of photons, but a purpose-made light source of sufficient beam intensity can also be used. Photons, being massless, travel at the maximum possible speed. Thus they are theoretically the highest possible exhaust velocity. By the same token, being massless, they do not transfer much momentum.

A. Photon Sails

As noted above, light does not transfer much momentum. So the following methods reach useful thrust level by maximizing area to mass ratio. The large area increases thrust relative to mass.The resulting structure resembles a terrestrial sail.

66 Solar Sail

Alternate Names: Lightsail

Type: Solar Flux by Photon Deflection

Description: Starlight, for example from the Sun, reflecting off a large area sail produces a force because the momentum of the photons is reversed. The force for perpendicular (normal) reflection is

$$(1+r)(E/c) ,$$

where r is the reflectivity of the sail, E is the incident light power, and c is the speed of light. Thus even perfectly absorbing surfaces see light pressure, but reflecting ones perform up to twice as well. No real surface is perfectly reflecting, so analysis of sails needs to take into account the actual optical properties. At sail angles other than perpendicular to the light source the net force has a direct component from the incident light, and a reflected or diffuse component depending on the surface properties of the sail. For reflection the incident and reflected angles are the same amount relative to the plane of reflection. The diffuse component is light scattered by a rough surface into a wide range of angles.The force is reduces by a cosine factor due to the spread of directions.

The absorbed component of light is eventually re-radiated by blackbody radiation. If this happens equally from both sides of the sail, it produces no net thrust. Coatings can modify the emissivity of each side and therefore give a net thrust. For high performance sails coatings add too much mass. It ends up being better to use bare metal foil of high reflectivity and just ignore the small part not reflected. Thickness so low as to be partially transparent, and micro-perforations smaller than typical light wavelengths can lower weight even further without significantly reducing reflectivity. Manufacturing, installation, and operation with extremely thin foil sails will be a challenge. The best mass-to-reflectivity material for visible light is an Aluminum/Magnesium alloy. Despite higher mass, refractory metal foils like Tungsten can be used at much higher temperatures. Therefore they can be used much closer to stars where light intensity is higher, and develop more thrust. This can be used for initial kick to outer system missions. The sail shape can be maintained by struts and tension wires, or by the balance of rotation and light pressure, or some combination.

At the distance of the Earth from the Sun, the incident power is 1370 MW per square kilometer. This produces about 8 Newtons/square kilometer for high-reflectivity sails. The attractiveness of solar sails is they use no fuel, and in principle the reflector can be made of local metals in the case of asteroids. The disadvantages are low thrust unless you build them very large, limited angular control, and inverse square dependence of thrust with distance from the source star

Status: As of 2012 a few test flights of sails have been attempted, with varying success. Any object exposed to sunlight will see a light pressure force, but most spacecraft do not have a low enough mass per area for the force to be more than a minor correction to it's motion.

Variations:

> **66a Gravity Tractor** - In this variation, the sail is not attached to a cargo such as an asteroid you are trying to move. The mutual gravity of the objects anchors the sail. Reasons to do this are not having to devise methods to attach the sail, and allowing the sail to maneuver itself with less restrictions.

References:

- Wikipedia article - **Solar sail**
- Marchal, C., **Solar Sails and the ARSAT Satellite - Scientific Applications and Techniques**, L'Aeronautique et L'Astronautique, no 127, pp 53-7, 1987.
- Friedman, Louis, **Starsailing: Solar Sails and Interstellar Travel** , Wiley, New York, 1988, 146 pp.

67 Laser Lightsail

Alternate Names:

Type: Laser by Photon Deflection

Description: A high powered laser is aimed at a target sail. The beam of photons are reflected off the sail material. Reflection of the photons reverses their momentum vectors' component which is normal to the sail. By conservation law, the sail gains momentum. Laser sails can have higher performance than solar sails because the laser beam intensity is not as limited as the local star light, and can be focused over longer distances. The sail can be designed to optimize reflectivity at the laser wavelength. This method is still by the overheating of the sail and lack of sufficiently powerful lasers for useful missions.

Advanced interstellar mission concepts have proposed very large phase plate type lenses for long range focus. Another advanced concept is to use two sails to slow down a vehicle, by releasing the first sail and using that to reflect light back to the second, smaller one.

Status: Not tested as of 2012 mainly due to lack of powerful lasers. Megawatt class lasers are not powerful in this context.

Variations:

References:

- Forward, Robert L., **Roundtrip Interstellar Travel Using Laser-Pushed Lightsails** Journal of Spacecraft and Rockets , vol. 21, pp. 187-95, Jan.-Feb. 1984
- "LightSail Project Documents"
- "Laser-pushed Lightsail"
- Fresnel plate lens for spacecraft propulsion. Gregory L. Matloff. "Deep Space Probes: To the Outer Solar System and Beyond". Section 7.5 The Fresnel Lens: A method of improving laser collimation. p. 103.
- "Laser Technologies for Starflight", "Key Issues for Interstellar Sails".

68 Microwave Sail

Alternate Names: Starwisp

Type: Beamed Microwave by Photon Deflection

Description: In this method microwaves are reflected off a very thin, open wire mesh. The momentum change of microwave photons bouncing off the mesh provides thrust. Because an open mesh of thin wires can have a very low weight, in theory this propulsion method can give high accelerations. With feasible power levels, the cargo mass will still be small. A good use for such would be delivering nanofactories, which can then build larger scale infrastructure. The beam can supply power to operate the factory at the destination. This is related to the non-transport energy delivery by microwave concept.

Status: Generating lots of microwave power is well understood. Keeping it focused over useful distances in space, and building the lightweight sail material are not.

Variations:

References:

- Wikipedia article - **Starwisp**

B. Photon Rockets

This group of methods involve emitting photons from inside the vehicle, rather than reflecting photons from an external source. Like photon reflection, photons do not carry much momentum, so it is a low thrust set of concepts, although the exhaust velocity is the maximum possible. It can be considered an addition to other propulsion methods by directional emission of radiator waste heat.

69 Thermal Photon Reflector

Alternate Names: Nuclear Photonic Rocket

Type: Nuclear Reactor via Photon Emission

Description: A heat generating device, such as a nuclear reactor, is at the focus of a parabolic reflector. The thermal photons are focused into a near parallel beam, which propels the vehicle. Another high-energy source is a matter-antimatter reaction, which is absorbed by a blanket of heavy metals and converted to heat. This method is probably not practical, since emitting photons gives less momentum than emitting the reaction products of the energy source directly. It might be worth using directional radiators to get rid of surplus heat in addition to the reaction productsIn that case it increases the total eficiency.

Status: Not tested as of 2012

Variations:

References:

- Wikipedia Article: **Nuclear photonic rocket**

70 Stellar Photon Engine

Alternate Names: Stardrive

Type: Solar Flux by Photon Emission

Description: This is similar to the previous Thermal Photon Reflector method, except an entire star is at the focus of a cloud of lightsails which are balanced by gravity vs light pressure. Since both gravity and light intensity vary as the inverse square of distance from the star, a specific thickness of sail will be balanced at any distance. The ratio of the star's luminosity to it's mass will determine the thickness. The star's spectrum and desired sail distance will determine what to make the sail out of. By directing the light in one particular direction rather than symmetric in all directions as natural stars do, the sails convert the star into an unbalanced light emitter, and so create a net thrust for the combined star + sail combination.

Sails with a slight dihedral angle oriented perpendicular to the star will direct the light to miss the star itself, and will be stable in orientation. Flat sails which do not reflect the light directly back to the star are slightly more efficient since none will be absorbed by the star, but will accelerate themselves sideways. So that type of sail must change it's orientation periodically to maintain position. Sails which are not flat, such as a shallow cone, can direct the light to miss the star without drifting.

To give a numerical example, a sail at the Earth's distance from the Sun produces about 8 newtons/km^2. Solar acceleration at that distance is 0.006 m/s^2.Therefore if your sailmass is 1330 kg/km^2, light pressure will balance gravity. If made of aluminum, then it needs to be 0.5 microns thick. Foils that thick are commercially sold today. A sail cloud reflecting 10% of the Sun's output will have an area of 28 x 10^15 km^2 and a mass of 37.5 x 10^15 tons (3.7% of the mass of the largest asteroid, 1 Ceres), Such a sail cloud would produce a thrust of 225 x 10^15 Newtons, and accelerate the Sun by 3.5 meters per second per million years.

Uses for this type of engine are adjusting orbits of binary or multiple star systems, escaping future supernovae, and generally moving stars within a galaxy or between galaxies. The physics of this method are simple. The challenge is one of scale, and the extremely slow accelerations it produces because stars are very massive. Note that moving a star does not generally bring along objects in orbit about the star, they would need their own propulsion.

> **Moving Planets -** A more practical (relatively) use for this method is moving planets, because of their much lower mass. The sails would be anchored by the planet's rather than the star's gravity. A larger secondary cloud of sails would direct more sunlight to the anchored sails than they would collect on their own. A reason to do this for the Earth is the Sun is increasing it's brightness by 10% per billion years due to increasing fusion rate in the core. Thus if you want to keep the Earth habitable, you would want to move it slowly outwards. Alternately, if you wanted to make Venus more habitable, you could move it into the Asteroid belt, where the excess greenhouse effect would be an advantage rather than a detriment. This would still likely be a slow method. Using gravity assist maneuvers with large asteroids or outer Solar System bodies would be faster. Unlike gravity assist for spacecraft, in this case the object doing the flyby is massive enough to affect the orbit of the planet. If a secondary body such as a gas giant is used to absorb the orbit changes, the flyby object could be used multiple times to shift the small planet without expending too much energy.

Status: Very far from testing.Possibly not even used in science fiction stories.

Variations:

References:

71 Gamma Ray Thruster

Alternate Names: redshift rocket

Type: Antimatter by Photon Emission

Description: Gamma rays produced by antimatter annihilation behind a vehicle can be absorbed by a thick layer of heavy metals. The momentum of the gamma ray photons produces thrust. Like other photon engines, it is likely more effective to use the decay particles directly, since they have more mass and can be directed better, but using the gamma rays as a supplement makes sense.

Status: Currently theoretical, as antimatter is only produced in tiny amounts in particle accelerators.

Variations:

References:

C. Amplified Photon Pressure

In this approach, a reflecting cavity is used to multiply the photon pressure. In one arrangement, a laser emits a particular frequency, and the walls of an enclosing cavity and the vehicle back end are designed to have a high reflectivity for that frequency. In another arrangement, there is no enclosing cavity, and a laser gain medium is placed in front of one mirror. The vehicle carries a second mirror, and the pair of mirrors plus gain medium form a laser amplifier which happens to have a large vacuum gap between the energy source and vehicle. In both cases the light will bounce multiple times until finally absorbed or dispersed, each bounce adding a momentum increment. The wavelength can be adapted to the physical dimensions, for example using microwaves instead of a laser if it creates an efficient reflection. The multiple reflections will extract energy from the photons as they repeatedly bounce off the moving vehicle by redshift. If sufficient bounces at sufficient projectile velocity happen, the light would be redshifted to a wavelength not efficiently reflected, or merely lose the majority of their energy, thus ending the momentum transfer.

References:

- Young K. Bae Corporation, **List of Articles**, 2006-2012. The paper on **Ultrahigh Precision Formation Flying** has a good illustration of the concept.

Section 2.9 - External Interaction Methods

External interactions are transport methods which create a force via fields or physical interaction with objects outside the vehicle itself.

Magnetic Interactions

72 Ionospheric Current Engine

Alternate Names: Electrodynamic Engine, Alfven Engine

Type: Electrical via Magnetic Field

Description: A current-carrying wire in a planetary magnetic field will experience an IxB force (current cross product with magnetic field, which is perpendicular to both). This is the same type of force created in electric motors. Currents which only flow in one direction cannot be sustained, electric charges would accumulate at one end. Therefore for this method the current loop is closed through an **Ionosphere**, the portion of a body's atmosphere which is ionized and can carry a current. The wire accelerates in one direction (pulling a vehicle along), and the ionosphere accelerates in the other direction. The ionosphere motion eventually dissipates within itself. When current is driven through the wire by a power supply, this functions as a motor, generating positive thrust. When current is allowed to flow in the other direction unopposed, it functions as a generator, producing power which can be used on the vehicle, and creating drag. Both power and drag can be useful in come circumstances, but any power generated will be at the expense of orbital kinetic energy.

A current loop thruster produces more thrust/watt than an ion or plasma engine. No propellant is consumed directly, although some material is consumed by a **Plasma Contactor** to produce a plasma that enables good electrical contact with the ionosphere. Effectively this gives an equivalent exhaust velocity of 250 km/s. Advantages are this is a relatively simple device, and relatively high thrust for an electric thruster. Limitations of this method are the planet must have both a substantial magnetic field and sufficient ion density to operate effectively, and the direction of thrust is governed by the local magnetic field direction. Planetary magnetic fields are often offset and tilted with respect to their poles, and in any case only point in one direction at any given point in orbit. The thrust direction is restricted to a circle perpendicular to the field direction. Earth orbit up to around 1000 km, and some areas around Jupiter are good candidates, other places less so. Uses include orbit maintenance for low orbit vehicles and Skyhooks, and apogee raising for reaching higher orbits. This method may be combined with other electric thrusters and share the same power source, changing which thrust method is used as appropriate for altitude or thrust direction.

The motion of a long wire in a magnetic field will generate a static voltage even when not used for propulsion. Such voltage build-ups happen to all spacecraft, a topic known as **Spacecraft Charging**, but are much larger with a long conducting wire. Therefore proper insulation and equipment grounding is needed to avoid damage. Even without the plasma contactor running, spark discharges to the surrounding ionosphere can happen.

Status: An experiment called the Tethered Satellite System was flown on the Space Shuttle in the 1980's.

Variations:

References:

- Belcher, J. W. "The Jupiter-Io Connection: an Alfven Engine in Space", Science vol 238 no 4824 pp 170-6, 9 Oct 1987.

Gravity Interactions

73 Gravity Assist

Alternate Names: Planetary Flyby, Celestial Billiards

Type: Kinetic Energy by Gravity Field

Description: Momentum exchange between a planetary body or large satellite and a vehicle allows changing the vehicle's direction and velocity in other reference frames. When considering just the vehicle and large body, a hyperbolic orbit has the same velocity on approach and departure, but gravity forces change the direction. When the large body is in turn in orbit around another larger planet or star, the change in vehicle direction can increase or decrease it's velocity relative to the larger velocity. By conservation of momentum, such gravity assist flybys affect the object you fly past, but typically the vehicle is so much smaller than the body it is flying past, so the change in the body's orbit is too small to measure.

The amount of velocity change is theoretically limited to twice the escape velocity from the body, by changing the vehicle direction 180 degrees. Thus larger bodies can produce larger velocity changes. As a practical matter gravity assist usually results in significantly less than this because the vehicle arrives with excess velocity and the desired final direction limits the flyby parameters. Excess velocity reduces the time the gravity of the body can act. More typical values are 0.5 to 1.0 times escape velocity measured from 1.5 body radii from it's center. The velocity changes are still large enough that they are extensively used in planetary exploration missions, often using multiple flybys of different planets, or even the same planet multiple times. The fuel cost of lining up for a gravity assist in these cases is much lower than doing the equivalent velocity change directly. The disadvantage is longer mission times consumed by the gravity assists compared to a direct transfer orbit, and needing to select mission dates according to when the planets are in the right positions. Lunar gravity assist to escape from or return to Earth orbit is particularly useful, as the Moon's orbit is short enough to give frequent opportunities.

Status: Used extensively in planetary missions, often multiple times for a single spacecraft.

Variations:

References:

74 Gravity Counterweight

Alternate Names: Dumb-Waiter System

Type: Potential Energy by Gravity Field

Description: In this method mass falling down a gravity well can be an energy source to power payloads going up the gravity well. Most typically this would be via a space elevator, using the falling mass to directly lift a cargo via cable, or to generate power to lift the cargo electrically. This is most efficient when cargo is going in both directions and are the same mass. In that case cargo delivery only consumes the inefficiency of the motors, which can be just a few percent. Another variation is to use braking energy of cargo above synchronous altitude, which sees a centrifugal force higher than gravity, to power lifting cargo from the body surface to synchronous altitude. The energy for this comes at the expense of rotation of the body by slowing it down. Braking will induce sideways forces on the elevator

Status: Counterweights are commonly used in elevators on Earth. Space applications have not been tried.

Variations:

References:

Aerodynamic Interactions

Interaction with the atmosphere of large bodies which have one, such as the Earth, can produce significant forces. These forces are distinct from propulsion using an atmosphere which was discussed under **Air-Breathing Engines**, which are principally thrust forces.

75 Aerodynamic Forces

Alternate Names: Aerobrake, Airfoils

Type: Kinetic Energy by Aerodynamic Forces

Description: This method uses aerodynamic interactions with an atmosphere to provide forces perpendicular to forward motion (lift), intentionally oppose forward motion (drag) or change orbit parameters (using lift, with some drag). Buoyancy, or lift by displacement without forward motion, is covered by method 3 Aerostat. There is a wide range of conditions and applications for aerodynamic forces and it is a well developed field of engineering.

The range of velocities is from well below the local speed of sound, or **Subsonic**, to many times the speed of sound, or **Hypersonic**, even beyond escape velocities. All bodies which are not symmetric or pointing directly parallel with the velocity direction will generate some lift, and all moving bodies generate drag. The ratio of lift to drag, known as **L/D Ratio** ranges from less than 1.0 for blunt re-entry bodies to 25 or more for good subsonic **Airfoil** shapes designed to maximize lift. Structures whose main purpose is to generate lift are called **Wings**. All lifting bodies induce some drag by virtue of the lift force not being entirely perpendicular, as well as body drag from other parts of the vehicle besides the wings.

Lift forces can be used to gain altitude as part of a climb to orbit. Drag forces would be minimized as they counteract thrust to reach orbital velocity. On return from orbit, purposely designed devices to slow down include parachutes, heat shields, and High Q Aerobrakes. While already in orbit, pure drag generating Low Q aerobraking can be done without special devices as long as the drag and heating are within what the structure can withstand. Orbital speed lift devices can change orbit direction, at the cost of some drag loss.

Status: Powered aircraft have operated for over 100 years. Jet aircraft have been used as the carriers for rockets. They use a combination of 25 Fanjet for forward thrust and wings for lift.

Variations:

- **75a Parachute** - Relatively low Mach number drag device to come to a complete stop or low enough velocity for a terminal landing device.
- **75b High Q Aerobrake** - Q is dynamic pressure, the pressure caused by motion through an atmosphere. High Q aerobrakes generate a lot of drag at high Mach numbers. They are often designed as inflatable devices, called **Ballutes** (from balloon and parachute), but extendable flaps or panels would also fall into this category.
- **75c Heat Shields** - They are drag devices integrated into a vehicle structure, but primarily designed to dissipate the extreme heat of re-entry
- **75d Low Q Aerobrake** - These are often re-purposed existing parts of a vehicle which are used for braking. By keeping the braking drag low, and using multiple orbits, often the braking can be done without special components on the spacecraft. Obviously that will take longer

References:

Mechanical Interactions

This category involves direct physical contact with a natural or human-made body in space.

76 Rheobrake

Alternate Names: Lithobrake, Crashportation

Type: Lower Kinetic Energy by Friction

Description: This methods uses mechanical friction against a planetary surface to slow down. For example, imagine a rail made of cast basalt on the lunar surface. It is laid level to the ground, and is shaped like the concrete highway dividers. A vehicle wanting to land is in a low grazing orbit. It aligns with the rail, just above it, then extends some clamps over the rail. By applying clamping pressure, the vehicle can brake from lunar orbit to a stop. Obviously the brake will be dissipating a lot of heat, and will therefore have to be made of high temperature material such as graphite. Another approach is to have a 'runway' which is a smoothed area on the lunar surface. The arriving vehicle slows down to below orbital speed, then gravity pulls it down to the runway, and friction with skids on the bottom of the vehicle slows it down.

In order to not melt the brake, it should maximize surface area and possibly have cooling systems like heat pipes. The rail will not see such concentrated heating, so cooling is not as much of a challenge. The main advantage is not requiring fuel to land on a body. The main disadvantage is the large size and mechanical accuracy required for the landing rail or runway. When the velocity is greater than the speed of sound in the brake materials, irregularities in the landing surface can create shock waves in the device. This method is easier to implement on smaller bodies where the amount of orbital kinetic energy to dissipate is less.

Status: Mechanical brakes are used in many vehicles on Earth. High performance ones are used in passenger aircraft. Use for space transport is theoretical at this point.

Variations:

- **76a Rock Cloud** - This involves creating an artificial 'atmosphere' of particles to slow down against. A cloud of lunar dust could be raised by electrostatic forces and an arriving vehicle slows by impact of the dust particles, or deflection by charged surfaces on the vehicle.

References:

76a Space Harpoon

Description: In 2015, NASA proposed using a harpoon and cable to hook to a passing object, and use braking of the cable to accelerate and match velocity. Redeploying the cable and using the objects spin to sling the vehicle would add additional velocity

---

Section 2.10 - Theoretical Methods

The previous sections included transport methods for which there is a scientific or engineering consensus they are possible to build, even if not built yet. For completeness, this section includes theoretical transport methods for which: a consensus does **not** exist, there is no known method to implement it, or even contradicts established physics. They are sorted alphabetically, since there is no reliable way to rank or organize these methods. Although there is not a firm technical basis for these methods, to be listed here, they need at least some theoretical support in the form of published papers or other documentation. Ideas which only appear in fictional works or have no theoretical support can be found in **Appendix 1: Fictional Methods**

77 Alternate Spacetime

Alternate Names: - Subspace, Hyperspace, Alternate Dimensions

Type: Theoretical

Description: - **Spacetime** in relativity theory is the 4 dimensional environment of three physical dimensions and one time dimension Travel in ordinary spacetime is limited to the speed of light as far as we know. This method uses the idea that there is some other spacetime which can be reached from ours. If it has different properties than ours, it could allow transport or communication faster or more efficiently. If an alternate spacetime is in relative motion to ours, which is not constrained by the speed of light, then rapid travel could be possible by translating to it, and then translating back at the destination. There are theories that our Universe actually consists of more than 4 dimensions, such as string or M-theory, but the other dimensions are compacted to the quantum scale, or unreachable. As of 2012, there is no evidence for these theories, although a considerable number of scientific papers have been written on the subject, and some searches for observable effects are underway.

Status: Theoretical as of 2012

Variations

References

Drosher and Houser, **Space Propulsion Device Based on Heim's Quantum Theory**, 2004. AIAA Paper 2004-3700. Assumes an extension of General Relativity using quantized higher dimensional space.

78 Antigravity

Alternate Names:

Type: Theoretical

Description: Antigravity is the reduction or opposition to the normal force of gravity, which is attractive under most conditions. One method of producing it would be with a negative mass. If such existed, the formula for gravitational force would produce a repulsion rather than attraction. No material with negative mass is known to exist. By Einstein's mass-energy relation ($E=mc^2$), negative mass would also represent negative energy. Other methods of producing repulsion have been proposed, but suffer a similar lack of observable support except for one item - **Dark Energy**. The Universe as a whole appears to be expanding at an accelerating rate. The cause of this is hypothesized to be a cosmological constant, a pressure that exists throughout the Universe tending to expand spacetime. Since there is no known way to change the pressure caused by Dark Energy, which is distributed evenly in all directions, it is not useful as a transport method.

Status: Theoretical as of 2012

Variations

References

79 Modified Newtonian Dynamics (MOND)

Alternate Names:

Type: Theoretical

Description: This method assumes some violation of Newton's laws of motion are possible. Either an action without an equal and opposite reaction, which produces a **reactionless thruster**, or higher order terms in the motion equations that would allow an unbalanced force. While such formulas are easy to write, they do not have support from actual observations. A resonant extraction of Casimir forces from the quantum background has been proposed as a way to produce thrust. While the Casimir force is well observed, using it in a way that generates reactionless thrust is not.

Status: Theoretical as of 2012

Variations

References

80 Quantum Black Hole Engine

Alternate Names:

Type: Theoretical

Description: In theory, all black holes will emit particles as if it were a black body of a certain temperature. This is known as **Hawking Radiation** after physicist Stephen Hawking, who first described it. The temperature varies inversely with the size of the event horizon, so smaller black holes are hotter and emit higher energy particles. The emission of a black body changes as the 4th power of temperature, while the area of an event horizon changes more slowly. Therefore small black holes emit more energy, and ones small enough to emit useful amounts of energy are themselves particle sized, and thus called **Quantum Black Holes**. If new matter is added to the black hole at a rate sufficient to offset the emission losses, effectively 100% conversion of matter to energy can be achieved. The particles or gamma rays thus emitted are directed for thrust or used for power generation. Black holes, quantum or otherwise, are very massive, so the utility of such for propulsion is questionable for anything smaller than an asteroid sized spaceship

Although there is a good amount of theory about quantum black holes, there is not a consensus that they actually exist beyond theory. The difficulty is in how to form sub-stellar mass holes. Holes can be manipulated by adding a net charge and then using electrostatic or magnetic fields.

Status: Stellar and larger mass highly condensed objects have been observed, and are presumed to be black holes. Quantum mass black holes have not been observed.

Variations:

References:

- Wikipedia article: **Hawking Radiation**

81 Tachyons

Alternate Names:

Type: Theoretical

Description: Tachyons are hypothesized particles which travel faster than the speed of light. They would either allow higher exhaust velocity for an engine, or by some sort of conversion, possibly by quantum tunneling, convert an entire vehicle into tachyons so it would travel faster than light.Some searches for tachyons have been made, but they have not been observed in nature.

Status: Theoretical as of 2012.

Variations

References

82 Warped Space

Alternate Names:

Type: Theoretical

Description: Travel through spacetime is restricted by current theory to the speed of light. Spacetime itself is not limited in this way, and in fact current **Inflationary Cosmology** theory assumes a faster than light expansion in the early history of the Universe. This method assumes that spacetime itself is distorted locally around a vehicle in such a way that apparent travel to outside observers is faster than to internal passengers. An example of such is the **Alcubierre drive**, but current theory does not indicate how to actually generate such a space warp. Note that in General Relativity theory, gravity is caused by a warp of spacetime, and passengers appear to themselves to not accelerate, while outsiders see them in accelerated motion. The difficulty is gravitational fields are not mobile, being attached to the large masses which cause them, so their utility for space transport is limited to methods like Gravity Assist, where you can make use of the diference in motion between two large objects.

Status: Theoretical as of 2012.

Variations

References

83 Wormholes

Alternate Names:

Type: Theoretical

Description: A **Wormhole** is a hypothetical region of spacetime shaped to connect two distant points. If the connection is shorter than the non-wormhole path, traversing it would save time. Creation of wormholes in theoretical papers usually involves black holes or **Exotic Matter**, matter with unusual properties such as negative mass. While many such papers have been written about wormholes, it is not known if the theory matches reality. Therefore we do not know whether wormholes are possible or what their properties might be.

Status: Theoretical as of 2012.

Variations

References

- Kanti, Panagiota et. al. Stable Lorentzian Wormholes in Dilatonic Einstein-Gauss-Bonnet Theory posted at arXiv.org 6 Dec 2011.

Section 2.11 - Comparison Among Methods

Transport methods can be sorted by performance, status, and cost measures. The following sections describe the measures and make a first attempt to rank the various methods. Technology is constantly progressing, so any selection of methods needs to be updated with the most current data, or projections for future planning purposes.

Comparisons Among Methods

Performance Measures

Exhaust Velocity

Any propulsion method that expels matter or energy to produce thrust obeys conservation of momentum, which is defined as mass times velocity. The exhaust momentum is equal to and opposite change in vehicle momentum. Given a finite supply of materials for the exhaust, mass is fixed. Therefore higher velocity leads to higher total momentum, and thus better performance for the vehicle. Exhaust velocity is measured in meters/second. Some systems, like 72 Ionospheric Current Engine, do not use exhaust for momentum exchange, but have a consumed supply - plasma generated for contact to the ionosphere in this case. The **effective exhaust velocity** is then calculated from the thrust force and mass flow rate.

Specific Impulse

Specific impulse is a derived measure traditionally used for launch from Earth, and then extended to space. It is defined as pounds thrust per pounds/second of fuel consumed, which simplifies to units of seconds. When multiplied by one Earth gravity (9.80665 m/s^2) it converts to exhaust velocity which is the preferred measure since it is in SI units.

Acceleration

Thrust to Weight Ratio

Status Measures

Technical Maturity

Implentation Status

Cost Measures

Development Cost

Production Cost

Operating Cost

Transport Optimization

Optimizing Vehicles

In past rockets, this has been done by using different type of fuel for different stages in a rocket. In the early part of the flight, air drag is important, so a dense fuel is preferred. A dense fuel means smaller fuel tanks, and hence less area to create drag. Thus the Saturn V used liquid oxygen/kerosene and the Shuttle uses solid rockets for the first stage, both being dense fuels. Both use liquid oxygen/ liquid hydrogen for the second stage. This has the highest performance in use for a chemical rocket fuel.

The Pegasus rocket uses an aircraft to get above the bulk of the atmosphere. A sub-sonic jet engine has about ten times the performance of a chemical rocket, mostly because it does not have to carry oxygen to burn.

Many, many propulsion combinations are possible in getting to Earth orbit and beyond. A large part of space propulsion design is choosing which methods to use and when to switch from one to another

Part 3: Engineering Methods

In Part 1 we described the fundamentals of physics and engineering that apply to any complex project. This includes Systems Engineering, which is concerned with managing the whole of a complex system across its entire life cycle, the design engineering tools and specialty areas of knowledge, and the organization and economics of projects. Finally we looked at what projects and programs already exist, and the categories of future projects which might be pursued. In Part 2 we began covering the particulars of space systems with the most characteristic element, the transport methods. Since humans are starting from the Earth, then transportation is a prerequisite to doing any other tasks in space, and so we discussed that first. Additionally, the very large number of possible transport methods justifies devoting a large section of the book to it.

In this Part 3 we will cover the particular design factors that apply to space systems, and the remaining subsystem elements besides transport methods. Combinations of subsystem elements then form complete end items or products which execute designed functions and missions. We will review these end functions and the methods available to perform them in a logical sequence by time, starting with exploration and ending with recycling. The final major part of the book with then treat combinations of multiple end items and systems - how they grow, interact, and evolve.

Overview

- **Design Factors**

Design factors are those which influence the whole of a design, across different subsystems. These include input requirements, technology level and availability of materials and suppliers, physical design such as margins and wear, and the limits imposed by humans as part of a system and the operating environment.

- **Subsystem Design**

There are numerous other systems besides propulsion required for most end items. These include structures, mechanical, power, thermal, data, communications, sensors, and environmental protection. When humans interact with an end item, you additionally need displays and controls, internal environment control, and crew support such as furniture, food, and clothing. Items with an extended life require maintenance and repair in the form of tools and spares, and supplies such as fuel.

- **Resource Exploration** - This includes the methods of finding and characterizing resources located in space, as far as their location, composition, and other physical properties. We only give a summary inventory of the known resources. Full details comprise the entire fields of space science and astronomy, which are both large and ever growing.

- **Resource Uses** - There are as many possible end uses for the available resources in space as there are things to do on Earth. We list the major applications in terms of what we want to do first. Later sections discuss how we can do them.

- **Resource Extraction** - Physical extraction is the task of removing materials from their native location, which is called **Mining**, along with preliminary processing and transport for further processing and production. Since you can only extract a physical material once from it's source, mining is typically mobile, where chemical processing and manufacturing tend to be fixed, since the equipment and connected power supply are often massive. Energy extraction involves converting a primary source of energy into more useful forms such as electricity.

- **Processing and Production** - These tasks combine a multitude of simpler operations into one or more complete process flows. The process flows convert extracted materials and energy into final bulk materials and finished components.

- **Assembly and Construction** - Assembly combines components to create a working device or machine. Construction first prepares a location, then assembles larger structures and outfits them with internal and external systems at fixed locations or orbits. The distinction between assembling mobile machines and fixed construction is somewhat arbitrary in space, since even the largest orbital constructions can be moved.

- **Operation and Maintenance** - Once items are assembled or constructed, they must be operated according to their intended use, and normally will require periodic maintenance to keep functioning. The overall concepts for operations and maintenance should be developed during design, so that necessary design features will be included. If a part will need replacing, for example, a way to access it and remove it should be part of the design.

- **Recycling Methods** - Most engineered items eventually reach the end of their useful life. Many methods and processes generate waste products, especially living things like humans. On Earth, disposal of scrap and waste has been left as a casual issue, and much of recycling, such as converting CO_2 back to Oxygen, happens by natural processes. In space, recycling has to be more deliberate. In many locations there are not surplus resources that can be used, you are floating in a vacuum. You also cannot just dump wastes, as they would end up being a debris hazard. Thus efficient recycling will be both necessary, and less expensive than extracting and delivering new resources in many cases. Thus recycling also needs to be planned for and designed in from the start.

While the topics arise more or less in this order, they need to be designed for in parallel, and will be performed in parallel, and in a connected network with transportation elements. The following pages will discuss these topics in more detail.

Section 3.1 - Design Factors

This section covers factors which influence a design across multiple types of subsystem elements. The most obvious influence is the set of requirements which define what the system is supposed to do and the criteria used to score how good the design is. Requirements and criteria were discussed in Section 1.5.

Technology Factors

These involve the performance levels of different technologies, and how mature they are. A project or design should assume consistent performance values and assumptions for a given technology. This needs to be done both across different subsystems within a design, and across different alternative design options. For example, if a solar panel power/mass ratio of 100W/kg is used in one alternative, the same value needs to be used in other design alternatives.

The maturity or **Readiness** of a technology is a measure of how far it has progressed from initial idea to commonplace use. NASA developed a scale of **Technology Readiness Levels** (TRLs) to describe the status of a technology. The higher the TRL, the less uncertainty about the cost and performance of the given technology. Care should be taken to understand the actual state of a technology, and allow appropriate risk margins.

Availability Factors

These involve whether a given material, component, subsystem, human skills, or facility will be available for a given project. A given design option might require components which are in limited supply. Even if they would perform better, if they are not available when the project needs them they will cause a delay which might be unacceptable. Reasons for limited supply include the number of suppliers with the skill and capacity to make the item, their current work backlog, upstream resources they need from their suppliers, import or export restrictions, possible production disruptions, intellectual property, and existing contractual agreements. Any item which is not obviously in abundant supply should be verified as to availability, to at least identifying a supplier who could meet the project needs. In some cases, the project itself can develop the capacity to provide a given item, but that imposes additional tasks to do so.

Physical Design Factors

- **Materials Selection** - This involves how to choose materials for different parts of the design. Usually materials selection involves multiple factors beyond the obvious ones like strength and melting point. Strength/density ratio is important when weight matters, which is most of the time for space systems. Materials cost. Qualification for a material mean sufficient testing to know how well a given material will work for a given purpose. New materials may not have sufficient testing.
- **Design Margin**
- **Design Life**
- **Corrosion and Fatigue**

Integration Factors

Integration is the process of combining lower level components into a higher level element which performs a set of functions.

- **Design Budgets**

A system will have budgets for finite items such as component mass, maintenance time, communications, or command inputs. These budgets apply across multiple subsystems, and must be estimated, allocated, and tracked.

- **Testability**
- **Subsystem Interactions**

This is consideration of the effects different systems impose on each other. These may include acceleration, vibration, thermal, electromagnetic, radiation, and others.

Human Factors

You cannot design humans (not yet, anyway), therefore you have to factor in human features into a design. This includes physical factors like acceleration tolerance, and mental limits such as the finite ability to learn and execute operational tasks.

Anthropometrics

Humans come in a range of sizes. Therefore seats and control interfaces have to accommodate different eye positions, eye focus, arm reach, hand size, and other characteristics. Devices such as space suits either need to fit the range of users, or crew selection needs to restrict the size range to fit the equipment. One source for data for space projects is the NASA **Man-Systems Integration Standards**, but this topic has gotten a lot of attention on Earth because most systems interact with humans, and thus have to be designed to interact with them.

Training

Operating a complex system is not intuitive, therefore human operators must be trained for the task. It imposes a requirement that the amount of natural skills plus training fit within what a given crew is capable of. For long duration operation, re-training may be needed for infrequent tasks. Simulators are used to do training at lower cost, and to train for hazardous conditions, such as loss of an engine. So in addition to designing operating hardware, training media, instructors, and simulators need to be considered in the design.

Environment Factors

Any design must be capable of withstanding the various environments in which it finds itself. These will include the production, storage, transport, and operating environments. We can divide environments in general into two groups, objects and space, since the former have distinct local conditions.

Object Environments

There are a wide variety of objects in the Solar System and beyond, of which the Earth is one. The environment conditions vary widely between them, and locally on individual objects, so we will not try to list individual details. Instead we will note the types of environment parameters which should be considered in a design, and what effects they may have. For specific locations, previous scientific or other data sources can be used for detailed information. If that is not available, new observations or visits may be needed to get the local data.

Atmosphere

If a body has an atmosphere, a number of conditions should be considered. These include:

- **Static Pressure** - Internal elements of the system, humans for example, may require particular conditions. If the outside pressure is too high or too low, then the design will need to include a pressure shell to maintain the desired level.
- **Dynamic Pressure** - These are the forces generated by a moving atmosphere (wind), or from moving the system through the atmosphere. The structure must be designed to withstand these forces in addition to any static pressure. Lightweight or large structures may bend or shift bodily from dynamic pressure loads.
- **Composition** - An atmosphere may contain gases which react with system hardware. Examples include oxygen and water on Earth, and sulfuric acid on Venus. Some gases are combustible (react quickly) with exhausted or leaked materials from the system. For example, a leaking oxygen tank may combust in a methane atmosphere.
- **Dust** - An atmosphere can transport dust and larger particles. These can abrade surfaces, or get deposited on them and accumulate. Dust may interfere with mechanical devices and be a hazard to living things.
- **Condensation** - Under certain conditions, atmosphere components can condense to liquid or solid form and then precipitate as rain and snow due to their higher density. Very small condensed particles may remain aloft as fog and clouds. The effects on a design include condensation on surfaces, and accumulation of fallen rain and snow on equipment and the ground.
- **Opacity** - An atmosphere can reduce visibility, communication, and filter incoming and outgoing light and heat due to being partially or totally opaque in particular wavelengths. The opacity can be caused by the gases, dust, or condensation within the atmosphere, and can be variable.

Temperature

The equilibrium temperature for a system element includes solar input, heat transport from any atmosphere present, and from the solid or liquid ground. Most objects rotate, so the Solar input will vary with time. If their orbit is significantly elliptical, the solar input will also vary with distance. Atmospheres can transport heat by radiation, convection, and conduction. To the extent an atmosphere is transparent or not present, heat can be lost to the very cold background temperature of the Universe. Systems operating close to, or on or below the surface of an object will get some heating from the object, in addition to solar input.

A functioning system also usually generates internal heat from operation of components, so the equilibrium temperature is what results when internal and outside thermal flows are in balance. If this temperature is higher or lower than desired for system operation, then components like radiators, heaters, and thermal insulation must be added to bring the internal temperature into the desired range

Gravity

On Earth, gravity level is within a few percent of the standard value (9.80665 m/s^2) for all locations. For other objects, except for the Gas Giants, it is generally lower and more variable. When the level is below biological or industrial process needs, then generating artificial gravity by rotation may be necessary. If the level is so low that traction or anchoring by weight does not work as it does on Earth, then special methods for movement and staying put may be required. When significant gravity levels are present, the design needs to account for structural loads caused by it, and bearing loads against the object, and ability of the object to support them.

Radiation

Almost all locations receive some radiation from a combination of radioactive decay, solar wind and flares, trapped particle belts, and cosmic rays. In addition to natural sources, a system may contain artificial sources such as radioisotope generators, accelerators, and reactors of various types. Humans and other biologicals, sensitive electronics, and some instruments are affected by high radiation levels. Protection comes from distance in the case of point sources, and shielding of various kinds from other sources. The best type of shielding varies by radiation type, and other parts of the design may provide shielding by their mass and arrangement.

Light Flux

The Sun is the major light source in the Solar System. It provides a source for heating and power, photosynthesis and other chemical reactions, and natural lighting. When not filtered by an atmosphere, the high energy part of the spectrum (ultraviolet and above) is a hazard to humans and other biologicals, and may degrade other materials. Lack of sunlight either from nighttime shadowing on an object, or inside system elements, may require artificial lighting.

Meteor Flux

On objects without a significant atmosphere, designs in exposed locations should account for the natural flux of meteors. Long term exposure causes pitting, and in rare cases larger impacts can cause more severe damage.

Object Changes

Object surfaces can experience transient events such as earthquakes, vulcanism, heavy precipitation and flooding, high winds, dust storms, fire, and others.They can also experience seasonal and long term changes such as surface melting and terrain shifts. A system should account for the frequency and severity of such events either by design features or by financial insurance against low frequency events.

Space Environments

The space environment does not have new classes of parameters which do not exist for large objects, but the details of each parameter will be different:

Plasma and Atomic Species

Significant atmospheres are bound by gravity to massive objects, but they shade imperceptibly into the background medium that exists between distinct objects. The upper reaches of the Earth's atmosphere extend past the lowest stable orbits. Although thin, the conducting plasma and atomic species in this region can affect hardware. This group of particles are distinct from the radiation group by having lower energy, and thus unable to penetrate solid objects.

Temperature

Due to the low density in the space environment it has low rates of heat conduction, and can therefore have very different internal temperature (defined by particle velocity) than that of objects embedded in it. For design purposes, system elements will mostly be affected by solar input, reflected light or shadowing by nearby objects, and heat loss to the cold cosmic background.

Gravity

Since gravity operates by an inverse square force law, it never vanishes entirely, merely decreasing in strength with distance. For objects in orbit, gravity forces manifest in the shape of the orbit, which influence design in ways like varying communications distance, sun and shadow times, and travel times to reach a desired destination. Free orbit trajectories for the different parts of a system are nearly identical when the hardware elements are small relative to the distance of a massive object. Therefore the design needs to account for the lack of forces between the elements. Large structures will see differences in gravity, called **Tides**, and objects using propulsion or rotation will see artificial forces that act like gravity

Radiation

Space environments typically have higher levels of radiation than found on the Earth's surface. The sources include UV and particle radiation from the Sun and cosmic rays. Bodies with strong magnetic fields, such as the Earth and Jupiter, can trap particles and create **Radiation Belts** with particularly high levels. Radiation levels can vary significantly from short term events like Solar flares. See the radiation heading under the previous Object Environments section for additional design factors from radiation.

Light Flux

Light flux in the space environment has the same design influences noted previously under Object Environments. The differences have to do with it being unfiltered by any atmosphere, and a different or lack of day/night cycle, depending on location.

Meteor and Debris Flux

The space environment contains natural solid particles ranging from dust grains up to whatever size distinct tracked objects are (nominally 1 meter). In addition to the natural particles, human-made debris, defunct hardware, and still functioning hardware also exist. Designs need to account for random impact of small particles, and tracking and avoiding larger objects, or otherwise accounting for the risk of damage from such impacts.

Section 3.2 - Subsystem Design

Elements by Subsystem

- **Structures**

  - Primary Structure
  - Pressure Vessels
  - Equipment Support

- **Mechanical**

  - Positioning and Translation
  - Access Hatches and Panels

- **Power and Electrical**
- **Propulsion**

  - Main Propulsion
  - Auxiliary Propulsion
  - Aerodynamics

- **Thermal**

  - Thermal Protection
  - Thermal Control

- **Data**

  - Processors
  - Storage
  - Network
  - Software

- **Communications**
- **Sensors**
- **Displays and Controls**

  - Ground Software
  - Data and Communications Hardware
  - Flight Software
  - Sensors and Instruments
  - Artificial Intelligence and Autonomous Operation
  - Trained Operators

- **Internal Environment**

  - Atmosphere Maintenance and Regeneration
  - Waste Management

- **External Environment**

  - Debris Protection
  - Radiation Protection

- **Crew Support**

  - Food
  - Clothing
  - Health Maintenance

- **Maintenance and Repair**

  - Tools and Equipment
  - Spares and Stocks

Section 3.3 - Resources: Exploration Methods

The energy required to move materials in space depends in large degree on the gravitational potentials that must be traversed. The deep gravity well of the Earth requires expending a large amount of energy to climb, as evidenced by conventional rockets. Therefore a primary way to reduce the difficulty and cost of space projects is to use local resources instead of bringing everything from Earth. To intelligently plan to use local resources, you first need to know what they are. The task of finding out what's there is called **Resource Exploration** both on Earth and in space. This section will discuss the methods of exploration, and a current inventory of known resources.The following sections will discuss uses for the resources and methods to implement them.

History and Process

All methods of finding resources involve first sensing the characteristics of a location, recording the measurements, and then interpreting the data to determine what it can be used for.Historically the sensing devices were the human organs of perception (eyes, ears, skin, taste, and smell). Paper and pen, and sometimes collected samples were used to record information, and then maps and written accounts made by humans interpreted the data. More recently, instruments with better accuracy and sensitivity than the human senses have been developed. Some instruments can detect properties that humans cannot sense at all. Integrating instruments with computers has automated the recording process, and is heading towards automation of the interpretation step. Humans are more in a supervisory role over the instruments and computers, though some amount of direct local sensing still goes on.

Scientific Exploration - Often measurements are made, such as wind speeds, for many years before a use, such as maps for wind turbine generators, were available. This goes to the distinction between scientific and commercial data. Scientific knowledge is sometimes described as good in and of itself. A less philosophical standpoint is that we do not know in advance what knowledge will prove useful. Therefore we accumulate all sorts of knowledge in the reasonable hope that some of it will. Space affects the Earth in many ways, and understanding the history of other planets helps us understand the particular one we live on better. Therefore astronomy and planetary science are expected to turn out useful in general.

Gathering data for "science" is also the first step towards more detailed exploration and local use, even if we don't yet know exactly what we will use it for. Without a definite use in mind, there is no benefit to keeping the data hidden or in duplicating effort, so science operates as a public and shared enterprise. That also provides better error detection and faster progress when more people can review and build upon past work of others.

Commercial Exploration - As uses become more local, private, and commercial, there is more of a tendency to keep data private rather than public. Individual gain and advantage now comes into play, and the relative effort of detailed exploration of one specific location is greater than general mapping or detection. So while similar equipment may be used for both public and private exploration, what gets done with the data is different. This is most evident in mining and other resource extraction industries. Some of the information gathered for commercial use will end up being released publicly and adding to the general store of knowledge. To date, almost all of the exploration of space has been done as science.

Progression by Distance

Exploration methods can be divided by the distance at which they are done, and the detail which they generate. These generally are inversely related - less distance generates more detail - because the instruments have a fixed sensor resolution. Greater distance from the interpreters of the data, as happens with unmanned planetary exploration, also has an inverse relation with transmission speed for the same reason. Devices like radio dishes to direct data back to Earth also have a fixed resolution, so greater distance makes for lower signal strength and bandwidth. So the advantage of smaller optics due to being close to the target has to be weighed against the need for more powerful transmitters and larger antennas due to being far from Earth.

The following sections discuss the various types of instruments classed into Long, Medium, and Short Range. Long range means the instrument is much closer to the storage and interpretation location than the target. Medium range means the instrument is much closer to the target than the storage/interpretation site, but not in immediate contact. Finally, short range means in immediate contact with the target location and able to interact with it.

Long Range Exploration (Astronomy)

Long Range Instruments

Telescopes:

Prior to 1610, long range exploration was limited to the human senses, and thus little was known besides the brightness and position of objects in the sky. The development of the telescope for astronomy by Galileo changed that situation, a change which continues to this day with larger telescopes and a broadening of the wavelength bands in which they operate. A telescope generally consists of optics which gather and focus low intensity photons (light, or electromagnetic waves), and a sensor which then detects the concentrated photons. This allows detection of what would otherwise be too dim to see. Originally the detector was the human eye, followed by photography starting in 1840, and by electronic sensors starting in 1979. Electronic sensors are up to 50 times more sensitive than film in terms of photon efficiency. A photographic plate might represent 400 megapixels of image resolution, so it is only recently that arrays of sensors also matched plate resolution as well as sensitivity.

Recording the data has progressed from hand written logs and charts of what the astronomer saw, to photographic plates which served as both the sensor and recording medium, to automated measurement of the plates to convert to digital form, to transfer of the electronic sensor data to computer storage. Interpretation of the data is now semi-automated with computer software. For example, detection of Near Earth Objects is made by comparing electronic images from two times. Anything that changes between the images is a candidate NEO, and the software filters known objects and variable stars, leaving a list of detections for a human to examine. Final interpretation of astronomical data is by humans, in the form of maps, catalogs, technical papers, and books.

Optical telescopes have grown in size, and thus sensitivity, migrated to better observing locations (high altitude and dry, or even in space themselves), and developed adaptive optics to get around the blurring of the Earth's atmosphere. For other wavelengths absorbed by the atmosphere, high altitude or space is a requirement. For long wavelengths, such as radio bands, single instruments do not provide much resolution. **Aperture Synthesis**, the mathematical combination of data from widely separated telescopes, is used to provide higher resolution in that case.

Wikipedia has several articles listing large instruments. **Optical Reflectors** are used for all the largest instruments because of the difficulties with large lenses. The active detector is usually an electronic Charge Coupled Device, or CCD, which converts light to digital signals, and often a spectrometer is used to sort the incoming light by wavelength. **Radio Telescopes** use a number of designs, with steerable metal dishes of various sizes being used often, and aperture synthesis by combining signals from different instruments up to the diameter of the Earth to get higher resolution. The active detectors are typically cryogenically cooled solid state amplifiers to reduce noise. **Space Telescopes** cover a number of wavelength ranges with a variety of detector types. The ones that operate near visible light wavelengths are designed similar to ground optical telescopes.

Gravity bends electromagnetic waves, and natural lenses formed by massive objects such as stars and galaxies have been used both to observe distant objects behind the lens, and detect and measure the lensing object. A conventional telescope is still needed near the Earth to use this effect. Such natural gravity lenses are not steerable. In the future the Sun may be used as a steerable gravity lens by placing the observing telescope at the right location opposite what you want to look at. The angle the Sun bends light requires a distance of greater than 550 AU to reach a focus. The attraction of using the Sun is the enormous diameter, which leads to high resolution. In the interim, other large telescopes of increasing power will continue to be built, and existing telescopes are upgraded with better instruments and continue to be used.

Meteorites

First some nomenclature, since it can be confusing: an asteroid or meteoroid are objects while still in space. A meteor is the bright trail in the atmosphere as it heats up and melts, and a meteorite is the object after it has hit the ground. Meteorites are direct long range samples of the space environment, where nature has brought them to us, instead of us having to go there and acquire samples. They provide useful comparison spectra in the laboratory, to compare to telescopic spectra from objects still in space. We can also analyze them directly to determine their composition. Meteorites are more useful for this purpose if we can record the arrival trajectory, which lets us guess where it started from. A drawback to meteorites is we have no control over where or when they fall. Sample return missions function in a similar way to meteorites as far as bringing the material to Earth for study at long range from the origin. They provide much better information on the source location and better preservation by returning the samples in a container rather than meteorite's open passage through the atmosphere followed by possibly long periods of exposure on the ground before being picked up.

Radar and Optical Ranging

Some objects get close enough to use radar or laser pulses to determine distance, and in some cases shape. The distance measurements are very accurate because of the timing accuracy of the detectors. Shape is obtained from time distribution of the return pulse, the parts of the object further away taking slightly longer. Several pulses at different times as the object rotates can be used to determine a three dimensional shape. Reflected pulse intensity varies as the inverse 4th power of distance, due to inverse square law both ways. So this method is strongly limited by distance, but the use of very large radio dishes and powerful transmitters has overcome that limit somewhat.

Long Range Data

The following types of information can be obtained by interpreting the long range instrument data:

Position/Orbit - Generally this is the first information found for a newly discovered object. The sky is mostly dark with stars and other objects showing up as bright spots to the sensors. Stars are slow moving relative to each other on human time scales, and the slower moving ones are referred to as the **Fixed Stars**, even though they are not if you wait long enough. For these types of stars their position in the sky is recorded in terms of latitude and right ascension projected onto a reference sphere assumed to be infinitely far away. Overlaid on the fixed stars are objects that move in short time scales. These are stars that exhibit parallax or orbital motion, and objects within our Solar System that show orbital motion. Parallax is a small shift in apparent position of relatively close stars vs farther stars caused by the motion of the Earth around the Sun. The width of the Earth's orbit, which is 2 Astronomical Units by definition, is small relative to the distance of even the nearest stars (260,000 AU). Nonetheless this allows direct determination of distance by simple trigonometry. All farther objects require estimating distance by indirect methods. Stars in binary or higher order systems (two or more stars), or ones with relatively heavy planets, can show a small motion caused by the collection of objects orbiting their common center of mass, rather than their own centers.

For objects within the Solar System, their orbital velocity is sufficient to show movement against the fixed stars in days or hours rather than months or years for stellar motions. Significant movement in a short time is the key feature to discover a Solar System object. Otherwise there is no way to distinguish it from the myriad stars which can be seen with a large telescope. By taking at least three measurements of position at known times, it's possible to determine the general parameters of an orbit. For discovery these are often at short intervals, such as successive nights. With additional measurements, and ones spaced out in time, which gives a longer baseline on the orbital path, the orbit parameters can be determined quite accurately

Size - The fraction of sunlight reflected from an object is called **Albedo**. Albedo times the area of the object gives the total amount of light reflected from it. From the orbit parameters you can calculate the physical distance. The observed brightness and an assumed albedo then allows an estimate of the size. On initial discovery the object may occupy only a single pixel on the sensor, so measuring size by the image on the detector can't be done, and estimate by brightness has to suffice.

Medium Range Exploration

Medium Range Instruments

Telescopes/Cameras

These operate the same way as long range telescopes on Earth, but due to being much closer to the target they get higher detail (measured as pixels/km or meter). Except for a few early film recorders, spacecraft have all used electronic sensors, whose data is usually recorded in a storage device, and then transmitted back to Earth. The storage is needed because the image recording time and when and how fast the transmission can be made are often different, especially when the telescope and transmitter are both fixed to the spacecraft body. Sensors designed for different wavelength bands can be used in the same instrument, and filters can be placed in front of the sensors to select specific wavelengths of interest. Farther infrared wavelengths measure thermal emission in addition to reflection of sunlight and the rate of change of a thermal map can indicate properties. Absorption of sunlight through an atmosphere can give it's composition.

Non-Imaging Optical Instruments

Besides direct 2D imaging via telescope optics, several other types of data can be collected:

- Radiometer - This measures infrared brightness to determine surface temperature.
- Polarimeter - Measures the polarization of incoming light by means of polarizing plates or films
- Photometer - Measures the total brightness of an object to high accuracy, without necessarily making a 2D image
- Spectrometer - Separates incoming light by wavelength, and records the brightness in each range. A large amount of information about an object can be determined from its spectrum.

Radar Instruments

Distance or altitude can be determined by the time a radar or laser pulse takes to bounce off the target. **Synthetic Aperture Radar** can determine altitude from the doppler frequency change vs return time.

Magnetometer

Measures the magnetic fields around a target.

High Energy Detectors

These instruments detect neutrons, alpha particles, ions, and gamma rays, which can provide composition information about a target. In some cases high energy natural radiation impacts the target, and secondary particles are emitted which are characteristic to the materials. In other cases particles are directly emitted by the target and detected.

Gravity Mapping

This is not an instrument in itself, but uses doppler and timing information from radio signals to infer motion of the spacecraft caused by the gravity field of the target. Two sensors in orbit can determine their separation very accurately by interferometry. Finer details of the gravity field can be inferred from changes in this distance.

Medium Range Data

Short Range Exploration

Short Range Instruments

Cameras

Cameras and telescopes are fundamentally similar devices, consisting of optics and an electronic sensor. The primary differences are at short range, large optics are not needed as much to get sufficient resolution or brightness, and targets are close enough to sometimes need focus closer than effective infinity. Two cameras separated by distance, or a single camera used from different positions can generate stereo data, from which three dimensional shapes can be inferred. When the optics are very short range and magnifying, the device is called a **Microscope**. Microscopic examination can determine mineral types.

Mössbauer spectrometer

This uses a gamma ray source, such as a radioactive isotope, and measures the recoil of atoms in the target to determine their composition.

Alpha Particle Spectrometer

This uses a combined alpha particle, proton, and X-ray source, such as a set of radioactive isotopes. The wavelengths and energies of the returned X-rays and particles are characteristic of the composition, which allows you to analyze a sample.

Laser Spectometer

A medium power laser is used to vaporize a target, and the resulting emission spectrum can be used to determine composition.

Chemical/Biological Sample Chambers

Samples of the target are collected and deposited in chambers, which are then subjected to various fluids and conditions such as heating. Various ways to analyze the result include: transmission spectra made by using a light on one side of the chamber and a spectrograph on the other, or gas chromatograph or mass spectrometer to determine composition of volatile gases that are released. Testing for biological activity from the target, or compatibility with Earth biology uses similar methods.

Kinetic Prospecting

Traversing the rough terrain of a body such as the Moon is difficult and slow. An alternative is send a lander/rover to a high point such as a mountain or crater rim.It picks up a rock of suitable mass and uses a centrifuge arm to throw it very fast at a selected target. Then it observes the impact with a telescope and spectrometer to determine the composition of the target, and possibly other instruments for additional data.Repeat as many times as needed for other targets. This allows prospecting a large area without having to drive over all of it.Kinetic methods have been used in the Deep Impact and LCROSS missions.

Ground Sensing Instruments

A **Seismometer** is very sensitive to motions of the ground. Natural or artificial impacts or movements generate seismic waves within the body. The timing of the waves as they arrive at the seismometer allows determining the internal structure and properties of the body. A **Heat Probe** inserted into the ground can measure total heat flow to and from the interior, and the thermal conductivity. A **Gravimeter** accurately measures the local acceleration of gravity from which density of the surrounding ground can be calculated.

Subsurface Examination

In **Core Sampling** a hollow drill is driven into the ground to separate a cylinder of nearly undisturbed material. The core sample is removed and examined above ground. **Drilling** uses larger machines to reach depths of up to several kilometers. The drill debris can be flushed to the surface and examined there, or instruments can ride behind the drill head or be lowered afterwards.

Short Range Data

Section 3.3 (page 2) - Resources: Inventory

Astronomy, and space exploration by governments, is a large field of research, and there is a lot of accumulated data about objects beyond the Earth. We will not attempt to reproduce all of that data here, but rather summarize and link to more detailed information. Many of the headings are links to Wikipedia or other detailed articles. Historically objects have been sorted as a hierarchy of Star > Planets > Moons, which is based on gravitational binding. Here we take the approach of sorting by available resources, mainly mass, composition, and energy, instead of gravitational dependency. The point of view is what is there that could be used, rather than the pure science approach of current condition and how did it form.

The inventory is organized by Solar System matter and energy, followed by our Milky Way Galaxy matter and energy, generally in order of decreasing quantities, and without regard to the practicality of using the resources. Resources outside our galaxy are too far away to be of near-term interest. Most of our Galaxy also falls into far-term interest, but some of the near portions may become useful given advances in technology, so we include it for completeness. Descriptions of large scale engineering projects may also involve extra-Solar resources.

Solar System Matter

History: The **Solar System**, which includes the Sun and all the objects gravitationally bound to it, formed approximately 4.57 billion years ago from a giant molecular cloud, likely along with other stars. Since it formed by gravitational collapse, the original composition of the cloud primarily determined the composition of the Solar System. Some material has since been lost by strong solar winds during early formation, and weaker solar wind to date, and by gravitational ejection of bodies. Additionally some Hydrogen has been converted to Helium by the Sun. The resulting composition found today is noted in the list of major elemental composition by object below. Aside from the Sun, most of the elements are combined into molecules and minerals which represent local minimal chemical energy.

The NASA Jet Propulsion Laboratory (JPL) **Solar System Dynamics** website provides information on most known natural bodies in the Solar System. Wikipedia also has an extensive **List by Size** of the larger objects.

The Sun

- **Mass**: 1.981 x 10^{30} kg, or 333,000 x Earth's mass. This is all but 0.14% of the total mass of the Solar System.
- **Mass Balance**: The Sun loses 4.28 million tons/second by virtue of the mass-energy of it's light output, and roughly 1.4 million tons/second due to the solar wind. It gains an (unknown amount) from comet and other object impacts.
- **Composition**: The visible surface of the Sun, known as the **photoshpere** consists of 74.9% H, 23.8 % He, and 1.3% heavier elements. Due to the relatively high gravity of the Sun, denser elements have tended to sink to the core. The bulk composition is estimated at 71.1% H, 27.4% He, and 1.5% heavier elements. From fusion of H to He, the process by which the Sun generates energy, the core is now roughly 60% He. Although the heavier elements are a small fraction of the Sun's composition, they still represent about ten times the mass of the rest of the Solar System besides the Sun.

The Gas Giants

Gas Giants are sufficiently massive to have retained a substantial fraction of Hydrogen and Helium, the lightest two elements. The Solar System has four, which between them have 444.6 x Earth's mass, or nearly all of the mass not included in the Sun.

Jupiter

- **Mass**: 1.899 x 10^{27} kg, or 317.8 x Earth's mass.
- **Mass Balance**: Jupiter's mass balance is not available, it would come from atmospheric loss and comet and asteroid impact gain.
- **Composition**: The atmosphere is about 75% Hydrogen, 24% Helium, and 1% heavier elements. The total composition is estimated as 71% Hydrogen, 24% Helium, and 5% heavier elements.

Saturn

- **Mass**: 5.685 x 10^{26} kg, or 95.15 x Earth's mass.
- **Mass Balance**: Not available.
- **Composition**: Saturn's atmosphere is 96.3% Hydrogen, 3.25% Helium, and 0.45% heavier elements. The total composition is estimated to be 20-32% heavier elements, and the remainder H and He.

Uranus

- **Mass**: 8.681 x 10^{25} kg, or 14.54 x Earth's mass.
- **Mass Balance**: Not available.
- **Composition**: Uranus' atmosphere is about 83% Hydrogen, 15% Helium, and 2% Methane. The total composition is estimated at 0.5-3.7 Earth masses rocky materials, 9.3-13.5 Earth masses ices (water, ammonia, and methane), and 0.5-1.5 Earth Masses H and He. The **ices** would be solids at this distance from the Sun, but in the interior of the planet they are actually hot dense fluids.

Neptune

- **Mass**: 1.024 x 10^{26} kg or 17.15 x Earth's mass.
- **Mass Balance**: Not available.
- **Composition**: The atmosphere is about 80% Hydrogen, 19% Helium, and 1.5% Methane. The total composition is estimated at 1.2 Earth masses rocky materials, 10-15 Earth masses ices, and 1-2 Earth masses H and He.

Objects With Atmospheres

Every object of sufficient mass and temperature will have some trapped molecules. For the purpose of this page, we define **Atmosphere** to be sufficiently dense to flow, rather than free molecule interactions, and is mostly non-ionized.

Venus

- **Mass**: 4.87 x 10^{24} kg, or 0.815 x Earth's mass. Atmosphere = 4.8 x 10^{20} kg.
- **Mass Balance**: Venus appears to be losing about 16 grams/sec x atomic weight of atmosphere due to solar wind stripping. Average accretion from comet and asteroid impacts is not available.
- **Composition**: The atmosphere is about 96.5% CO_2, 3.5% Nitrogen, with trace compounds. Because of similarity in density and total mass to the Earth, Venus is expected to have a similar total composition and structure, consisting of a rocky outer portion, and a metallic core.

Earth

- **Mass**: 5.974 x 10^{24} kg, or exactly 1.000 x Earth's mass. Atmosphere = 5.28 x 10^{18} kg.
- **Mass Balance**: Earth loses about 3 kg/s of Hydrogen and 0.05 kg/s of Helium from the upper atmosphere. A small amount of man-made objects depart the Earth per year. On the order of 1.25 kg/s of extraterrestrial material of all sizes (dust to dinosaur killing asteroids) accretes. Therefore as a whole the Earth is losing mass, but the current rate amounts to 40 parts per billion of the total mass over the remaining life of the Earth.
- **Composition**: The atmosphere contains 78% Nitrogen, 21% Oxygen, 1% Argon, variable amounts of water, and trace compounds. The total composition consists of the following major elements: Iron (32.1%), Oxygen (30.1%), Silicon (15.1%), Magnesium (13.9%), Sulfur (2.9%), Nickel (1.8%), Calcium (1.5%), and Aluminium (1.4%); with the remaining 1.2% consisting of trace amounts of other elements. Due to density segregation, the core is 95% Iron and Nickel, while the uppermost layer, the **Crust**, is nearly all metal oxides. By weight these are oxides of Si, Al, Fe, Ca, Mg, Na, K, H, C, Ti, and P.

Mars

- **Mass**: 6.42 x 10^{23} kg, or 0.107 x Earth's mass total. Atmosphere = 24.8 x 10^{15} kg.
- **Mass Balance**: Loss of Oxygen ions due to the solar wind is estimated at 16-35 grams/sec in a 2010 article by Fang et. al., **Effect of Martian Crustal Magnetic Field on Atmospheric Erosion**. At this rate the atmosphere would be eroded in 22.5 billion years. Accumulation from incoming asteroids and meteorites is not available.
- **Composition**: The atmosphere contains 95.3% CO_2, 2.7% Nitrogen, 1.6% Argon, and trace components. The total composition is generally a silicate outer layer with a metallic core, with elements by mass roughly in the order oxygen, silicon, iron, magnesium, aluminum, calcium, and potassium.

Titan

- **Mass**: 1.345 x 10^{23} kg, or 2.25% of Earth's mass in total. Atmosphere = 9.05 x 10^{18} kg.
- **Mass Balance**: Not available, but considerations are stripping by Saturn's magnetosphere, accretion from small objects, and stripping from larger impacts.
- **Composition**: The atmosphere is variable by altitude. Higher levels are 98.4% Nitrogen, 1.4% Methane, and 0.2% trace compounds. Lower levels are 95% Nitrogen, 4.9% Methane, and 0.1% trace compounds. Total composition is roughly half silicate core and half ices and likely liquid water at some depth. The surface appears to be mostly water ice, with variable amounts of Carbon Dioxide, Methane, Methane Hydrate, Ammonia, and Methanol by location.

Triton

- **Mass**: 2.14 x 10^{22} kg, or 0.36% of Earth's mass in total. Atmosphere ~5 x10^{13} kg.
- **Mass Balance**: Not available.
- **Composition**: The thin atmosphere is mostly Nitrogen. The solid composition based on density is likely 30-45% water ice, and the remainder rocky core. Surface is 55% solid Nitrogen, 15-35% water ice, and 10-20% carbon dioxide ice.

Objects Without Atmospheres

This group includes those without significant atmospheres. This happens whenever the combination of low mass and temperature allows gas molecules to escape or to be stripped off by Solar wind interaction. They are grouped by the International Astronomical Union into one planet (Mercury), several dwarf planets, the remaining satellites of the major planets, a very large number of **Small Solar System Bodies** which are further divided into types, and a possible class of **Rogue Objects**. These divisions into groups are more for convenience by orbit location than mass, composition, shape, and internal arrangement. There is a more or less continuous distribution of all of these properties, including orbits. Because most orbits are elliptical, and sometimes highly so, a single object can cross several nominal orbit locations. Wikipedia has a table of **Solar system rounded objects** which includes the dwarf planets, larger satellites, and dwarf planet candidates.

Mercury

- **Mass**: 3.3 x 10^{23} kg, or 5.5% of Earth's mass.
- **Mass Balance**: Not available.
- **Composition**: Approximately 70% metallic core and 30% silicate outer layers.

Dwarf Planets

Dwarf planets are defined as large enough to have been shaped by gravity (into an ellipsoid), but not so massive as to clear its neighborhood of other objects. The category was created in 2006 when it was obvious that Pluto was part of the Kuiper Belt region, which contains some objects larger than the former planet. Currently 5 dwarf planets are officially recognized (1 Ceres, Pluto, Haumea, Makemake, and Eris), and it is expected several hundred to 2000 others will be found as observations of the region beyond Neptune improve. Objects larger than 838 km in diameter are provisionally classed as dwarf planets, but final status requires they be rounded, and most such objects have not been observed that well yet.

- **Mass**: Five official dwarfs: 3.8 x 10^{22} kg, or 0.63% of Earth's mass. Perhaps 10% Earth mass once all outer system objects are discovered.

Planetary Satellites

This group of objects is distinguished by being attached to a major planet and not large enough to maintain a significant atmosphere. The Wikipedia **List of Natural Satellites** shows the known ones to date, from which Titan (Saturn VI) and Triton (Neptune I) are excluded since they have atmospheres and were listed above. It is very likely there are additional small satellites orbiting the major planets which have not been discovered yet. Total mass is around 475 x 10^{21} kg, mostly in the 4 largest moons of Jupiter and the Earth's moon. Composition varies by distance from the Sun from rocky (the Moon) to icy. We group these objects into two classes by size: Large and Small. The large group are shaped by gravity, and possibly internal heating from radioactive decay or tidal friction from the planet it orbits. This group are approximately spherical, and would be considered dwarf planets if not orbiting a planet themselves. The more numerous Small group are irregular in shape and likely not differentiated by internal melting.

Outer System Objects

The outer system includes objects outside the orbit of Jupiter, but not the outer planets themselves or their associated satellites. In approximate order of semi-major axis (symbol a, in AU) this includes the **Centaurs**, **Kuiper Belt**, **Scattered Disk'**, **Hills Cloud**, **Oort Cloud**, and **Rogue Objects**. There is overlap between these groups and with asteroids. The distinctions are based on convenience in describing their orbits and composition rather than a real difference in type. They are all objects not attached to a major planet with orbits beyond Jupiter and too small to maintain a permanent atmosphere. All the outer system objects are presumed to form from the same solar disk which formed the planets, but were then scattered by the gas giants once they grew massive enough. **Comets II** and **The Solar System Beyond Neptune** are recent reviews of the state of knowledge for some of these objects.

The total mass of all the outer system objects is poorly known and is an active area of study. Density and spectra indicate the outer system objects as a whole consists mostly of Hydrogen-bearing ices: Methane (CH_4), Ammonia (NH_3), and Water (OH_2), which are all solids at large distances from the Sun. Comets are former outer system objects whose orbit now comes closer than Jupiter to the Sun, so that gas evaporates of their surface, thus they are listed below. The remainder stay cold enough to be solid.

- **Centaurs**

These are objects which have closest approach to the Sun (perihelion) between Jupiter and Neptune, and no limit on the farthest point. Approximately 200 are known as of 2012, and are being discovered at about 15-20 per year. The discovered ones range in size from 200 down to 2 km in diameter. Estimated number larger than 1 km in size is 44,000. Because they cross the orbits of the gas giants, their orbits are unstable on a scale of a few million years.

- **Kuiper Belt**

The Kuiper Belt includes objects in stable orbits outside the orbit of Neptune (30 AU) up to 50 AU, having eccentricity generally from 0 to 0.2, except for resonant objects which can go to about 0.4, and with inclinations up to about 35 degrees. Pluto is now considered the largest such object, and 1241 such **Trans-Neptune Objects** (TNOs), as they are also known, have been found since 1992. Estimated number of TNOs larger than 100 km diameter is 0.1 million, >10 km possibly as high as 10 million. Estimated mass is 0.04-0.1 x Earth's mass, but models of planet formation predict there were 30 Earth masses of such objects originally. Where the rest went in that case is unknown at present. Composition in general is poorly known, but a combination of CH4 (methans), NH3 (ammonia), and H2O (water) ices are expected from the small number of density and spectral observations.

- **Scattered Disk'**

Scattered Disk Objects (SDOs) have perihelion (closest approach to the Sun) beyond Neptune (ie 30 AU) with higher eccentricity (0.20 to 0.94) at a given distance, and extending to larger distances than the Kuiper Belt. Their inclinations range up to about 40 degrees. The rise in eccentricity with distance may be a selection effect, since we can only find objects that come closer to the Sun. Outer system objects observed from Earth dim as the 4th power of distance. This is the product of solar intensity, which falls as the square of distance, and angular area, which also falls as the square of distance. The name comes from their orbits being gravitationally scattered by the gas giant planets at some point in their history from the lower inclination and eccentricity of the Solar nebula from which they formed. Their orbits are subject to further changes from planet interactions, so are not stable over the long term. Since 1995, a total of 167 **SDOs** have been found (in that table, the ones with q >30 AU). Their total mass is estimated at 0.01-0.1 Earth Mass.

- **Hills Cloud**

The Hills Cloud, also known as the Inner Oort Cloud, are objects with semi-major axes between 1,000 and 10,000 AU. The Sun is expected to have formed in a star cluster embedded in a gas cloud. **Computer simulations (2011)** of such clusters indicates objects forming closer to the Sun would be scattered out to these distances and into the outer Oort Cloud with an efficiency of ~1.5%, leading to an estimate of ~3 Earth masses currently. The Hills Cloud is bound strongly enough to the Sun that perturbations from outside sources and the gas giants rarely put them in orbits where we can detect them. Their large distance makes them too dim to see except when they approach closest to the Sun. Thus we have only discovered a few objects on orbits which enter this region.

- **Oort Cloud**

The Oort Cloud is made up of objects with semi-major axis > 10,000 AU. At such great distances from the Sun, their orbits are affected by close star passes and galactic tides. Their origin may partially be capture of loose objects in the cluster where the Sun formed. The remainder would be highly scattered from the solar disk during the Sun's formation. Estimated mass may be 0.1-7 Earth masses, consisting of about 6 x 10^{10} to 10^{12} objects. This estimate comes from observing the frequency of long period comets, but this does not place strong limits on their number.

- **Rogue Objects**

Rogue objects, also called nomads or rogue planets, are not bound to the Sun and merely passing nearby. None have been definitely identified in our vicinity, but in 2011 some candidates in our galaxy were discovered through gravitational microlensing. Since Solar System objects have been ejected by the gas giants, by symmetry there should be objects ejected from other stars which are currently near the Sun. In addition to ejected objects, objects that formed separate from stars by the same mechanism are expected. Total numbers are very uncertain, with estimates from 2 to 100,000 such objects per star, including our own.

Inner System Objects

This group includes objects from the orbit of Jupiter inwards. It is listed after the outer system objects because this is a resource-based inventory, and the total mass is smaller. In decreasing distance from the Sun it includes:

- **Jupiter Trojans**

There are stable **Lagrange Points** located 60 degrees ahead and behind a smaller body orbiting a more massive one. The most massive such pair in the Solar System, Jupiter and the Sun, has the strongest such stable regions, extending about 30 degrees of Jupiter's orbit each. Objects trapped there are called **Trojans** because the first few found were named after mythical characters from the Trojan War. As recently as 1961, only 14 Jupiter Trojans had been discovered, but by 2012 the Minor Planet Center lists over 5300. There are estimated to be 0.1-0.3 million Trojans larger than 2 km diameter. Total mass and composition are poorly known at present, but can very roughly be estimated at 2 x 10^{19} kg. A relatively small number of similar Trojan objects exist around other mass pairs.

- **Main Belt Asteroids**

About 600,000 total asteroids are known in the Main Belt which ranges from 1.3 AU minimum distance to Jupiter's orbit. The dense region of the Main Belt has a doughnut-shaped volume of about 70 AU^3, and so an separation of 0.05 AU, or 7,300,000 km. The number of known asteroids has increased by 60 times since 1980, and is expected to continue growing, but newly found ones are generally small. The total mass in this region is estimated at 3 x 10^{21} kg (0.05% of Earth). Total number larger than 1 km is roughly 1 million, and >100 meters ~25 million.

- **Near Earth Objects (NEOs)**

These are defined as all objects which have a closest approach to the Sun of less than 1.3 times the Earth's distance (1.3 AU), with th exception of the Sun itself, Mercury, and Venus. Mars has an average distance of 1.52 AU, but is lower mass than the Earth. The 1.3 AU definition is thus approximately closer to the Earth than outer planets from a gravity standpoint, or close enough to be interesting. This is an artificial definition from a human interest standpoint - there is no distinct physical grouping for this class of bodies, such as there is for the Main Belt asteroids. NASA has a Near Earth Object Program to discover and characterize them, and as of mid-2012 there are about 9000 known. Estimated number >1 km size is around 1000, and >100 m is about 200,000.

The NEO category includes all object types, including asteroids, comets, extinct comets (which can be hard to distinguish from asteroids), and manufactured spacecraft. There is an undefined lower bound to size, which we will assume to be 1 meter diameter, below which we refer to them as meteoroids, dust, or particles. The population of NEOs is not permanent. Over periods of approximately 1-10 million years either gravity effects from planets and larger bodies, or collisions with them, will remove them from the NEO orbit range. Objects smaller than approximately 1 cm are also affected by light pressure or other effects more than gravity interactions, and have even shorter lifetimes in those regions.

Comets

Comets differ from the solid objects listed previously in that they show periodic vaporization and dust emission. This is caused by heating of their surfaces when they get close to the Sun. Water sublimation becomes strong at less than 2.5-3 AU, and other volatiles at other distances. The lost material can create spectacular though not very massive tails. Since comets are in effect boiling gases from their surface, the built up pressure can cause them to fragment. Their lives are limited by the amount of volatiles they start with. Defunct comets resemble asteroids, and distinct trails of debris in their orbits are observed as periodic meteor showers. Comets ultimately originate in the Oort Cloud, and migrate to closer orbits by gravitational perturbations.

Comets are divided by their orbits into **Short Period**, **Jupiter Family**, **Halley Family**, and **Long Period** classes. Typical orbit changes in terms of semi-major axis are about 0.001 per orbit. If there were no planets around the Sun, the comet orbits would tend to stay fixed. Since Jupiter is roughly 1/1000 the mass of the Sun, it can be thought of as a blender blade mixing up the orbits by that amount per pass. Therefore comet orbits tend to randomly migrate among orbit classes.

Jupiter Family

Since Jupiter is by far the most massive planet in our Solar System, it has the most influence on comets which cross it's orbit. There is a noticeable cluster of comets whose maximum (aphelion) distance (Q) from the Sun is close to that of Jupiter (5.2 AU). The cluster roughly ranges from 4.2 AU < Q < 11 AU, with minimum (perihelion) distance (q) ranging from 0.5 AU < q < 5.5 AU. This range is generally just outside that of the main Asteroid Belt. Jupiter family comets generally have inclinations less than 35 degrees. Approximately 200 Jupiter Family comets are known.

Halley Family

This group of comets have orbits with periods of 20-200 years which are oriented so that Jupiter does not affect them strongly. Their inclinations range from 0 to 180 degrees, with a slight excess below 60 degrees.

Long Period

These are defined by having an orbital period > 200 years, and thus a semimajor axis (a) > 34 AU. Long period comets strongly cluster in semi-major axes near 10,000 AU, which makes them hard to distinguish from parabolic. This cluster of orbits led to the assumption of the Oort cloud as their source. Their inclinations span the full range from 0 to 180 degrees.

Particulates

Dividing line between distinct objects and regions or masses of particles better tracked as a whole.

Particle Belts

Rings

Interplanetary dust Gas and solar wind

Solar System Energy

The Sun

Current Energy Output From fusion of hydrogen to helium

Estimate Energy Reserves

- Hydrogen Fusion
- Additional Fusion Reactions
- Stored Thermal Energy
- Gravitational Collapse Energy
- Minor Energy Reserves - Spin, stratification, magnetic field

Everything Else

Everything else is lumped under one heading because magnitude of the Sun's energy reserves is so much larger than everything else combined.

Latent Heat of Formation - When massive objects like planets formed within the Solar System, the collision and gravity well energies, and later the stratification of the interiors by density released a lot of energy, part of which went to heating the interior of the object. For the larger objects, some of that heat is still stored in their interiors.

Nuclear Fission - Objects in the Solar System incorporated elements with radioactive isotopes which decay naturally, heating their interiors, or can be made to fission on purpose.

Nuclear Fusion - Just like the Sun, but on a smaller scale, there is potential energy in light elements that can be fused together. Since it takes at least 75 Jupiter masses for this to happen naturally, then it must be made to happen artificially.

Chemical Reactions - Materials such as fossil fuels and undecayed plant matter can burn with oxygen in the atmosphere to release energy

Orbital Kinetic Energy - The bulk of the kinetic energy in the Solar System resides in the motion of the planets and smaller bodies. When spacecraft use a flyby to alter their motion, they extract a little of this energy.

Minor Energy Reserves -

Galactic Matter

The Galaxy or Milky Way in capitalized form refers to the gravitationally bound object which the Sun and Earth orbit within. In lower case, **galaxy** refers to the general class of such objects.

Summary description of the Galaxy as a whole: total mass of the Galaxy, and baryonic vs Dark Matter)

Components sorted by Mass:

Dark Matter

Stars

Satellite Galaxies

Clusters

Clouds

Central Black Hole

Substellar Objects

Brown Dwarfs -

These are objects too small to count as stars, but above the limit for planetary bodies. The upper limit is about 80 times Jupiter's Mass, above which hydrogen fusion can happen and the object is considered a star. The lower bound is about 13 times Jupiter's mass, below which no fusion will occur. Above this lower limit, Deuterium and Lithium fusion can happen, but since these are much rarer than Hydrogen, it limits their life and brightness.

Planetary Systems

In recent years a large number of planetary systems around other stars have been detected. The **Extrasolar Planets Encyclopedia** catalogs them with references to original papers. As of 2012 there were 660 such systems with 837 planets detected. Methods of detection vary. In addition to planets which have formed, **Circumstellar Disks** represent early stages of formation or incomplete condensation to larger objects.

Rogue Objects

These are objects not tied to a star, which can range from planets below the brown dwarf limit down to ejected comets. They are currently not well understood, as only a few candidates have been **detected** by gravitational lensing in 2011. **Estimates** of their number range from two per main sequence star up to potentially 100,000 per star. The larger number depends on optimistic assumptions for the mass function (number of objects vs their mass).

Interstellar Medium

Dust Particles

Interstellar Gas

Galactic Energy

Total Power output

Total Energy reserves

Fusion

Gravitational energy

Dark Energy

Angular Momentum

Thermal Energy

Magnetic Fields

High Energy Sources

This includes cosmic ray flux and X-Ray sources.

- Wikipedia article: **Ultraluminous X-ray Source**

Retrieved from "https://en.wikibooks.org/w/index.php?title=Space_Transport_and_Engineering_Methods/Resource_Exploration2&oldid=3220346"

Section 3.4 - Resource Extraction

This section discusses how to extract resources from their natural state. For matter and energy these are commonly called mining and energy production, but the latter is a misnomer. Energy cannot be created, it can only be converted from an existing source, so we prefer the term **Extracted**.

Mining

Mining is the process of extracting, or in the case of dispersed materials, collecting, physical materials for direct use or further processing. A **Deposit** or **Ore** is a naturally occurring material of sufficient size or concentration to be mined economically. Mining techniques will vary according to what you are extracting and where you are doing it. So this section organizes the methods by physical state and type of natural body being mined. Some processing may occur at the mining site to concentrate the ore or prepare it for transport. Concentration is called **Beneficiation**, and there are a number of other processes which can be applied, such as crushing and sintering.

The Challenge

Most of the baryonic mass in the Universe is inconveniently located in the interior of large bodies, where it hard to get to. In fact spheres, the shape which many large objects approximate, have the least surface area for a given volume. In other words, the ratio of relatively inaccessible material in the interior to accessible material on the surface is a maximum. Another problem is that useful metals, such as Iron, tend to collect in the center of large objects due to their density, where they are inaccessible due to surrounding layers of rock. Besides physical inaccessibility, much of the matter is in a plasma state (e.g. in stars) or very low density (e.g. in molecular clouds). So the challenge for mining is accessing the small fraction of currently accessible material, and developing techniques to increase the range of such materials.

Even conventional mining on Earth literally only "scrapes the surface". So-called deep mining and drilling typically only reach the top 0.1% of the planet's radius. This limits our ability to obtain rare materials, or enough of common materials for future large projects. Mining in space is one method to increase the range of accessible materials, because it is new accessible area with different history and composition than Earth, but it should be considered a part of the total challenge for future mining.

Mining Solid Bodies

Mining by G-Level

We can divide solid bodies into groups by size. Small bodies are such that typical mining equipment forces and velocities are larger than local gravity. Therefore equipment anchoring to keep it from moving itself is needed, or special design to contain gravel and dust from traveling long distances, or even leaving the body entirely. By extrapolation from Earth construction equipment, local gravity below 2 m/s^2 (0.2 g) will require lower equipment forces or anchoring. Blasting is the most energetic mining operation. From blasting safety codes we estimate that gravity below 0.3 m/s^2 (0.03 g) would require safety distances of over 1 km from shrapnel and debris, which becomes unreasonable. That value will likely need updating from actual experience. For now we will divide bodies into low, medium, and high gravity for mining purposes at 0.3 and 2 meters/s^2. For the very smallest objects you do not mine them in the sense of removing material from a larger body, but capture and transport them to a processing location if needed.

Low-G Mining

Medium-G Mining

High-G Mining

Mining by Depth

We can also divide mining techniques by depth. These are surface, sub-surface or underground, and deep mining. Sub-surface is anywhere an open pit or angle of repose is not sufficient, and columns or walls are needed for support. The support can be part of the natural materials left in place, or artificial supports installed as part of the mine development. Deep mining is when the gravity loads on the natural materials from what is above it start to cause shifting or collapse. This requires fully enclosed tunnels or pipes or other special techniques. The strength of the surrounding material will determine when you reach a deep condition, and this will vary from place to place. As a guide, for rocky locations, deep can be considered more than 15km /(local acceleration of gravity in m/s^{2}). Gravity falls as you move to the center of a body, so aside from the few largest asteroids, the whole interior may not reach a deep condition from lack of sufficient gravity load. Conversely, on a larger body such as the Moon, you can reach a deep condition within 0.5% of its radius due to higher gravity and larger radius.

Surface Mining

Sub-Surface Mining

Deep Mining

We will take as an example of deep mining to excavate the core of the second most massive asteroid, 4 Vesta, for Iron. Vesta melted in its early formation from radioactive decay and formed an Iron core and rocky mantle. It has the following properties:

- Polar Radius = 229 km surface, 106 km core
- Equatorial Radius = 305 km high point, 114 km core
- Gravity = 0.329 polar, 0.1532 equatorial highpoint, and 0.257 core in m/s^2. The much lower equatorial gravity is due to the larger radius and rapid rotation (5.34 hour rotation period).
- Mantle Density = 3115 kg/m^3

As an approximation we take the polar and equatorial average acceleration from surface to core times the distance. This amounts to 123,000 m x 0.293 and 191,000 m x 0.205. The polar value is slightly lower, so we dig down from the pole. A column of mantle rock of that height, density, and acceleration sees a pressure at the bottom of 112.26 MPa (16,300 psi). This is higher than the likely strength of the rock, so an unsupported hole will likely fail, and we are indeed in a deep mining condition. Therefore the hole will need a lining such as steel obtained elsewhere, at least for the deeper parts. The core itself, being made of mostly Iron, will be self-supporting until too much of it has been extracted and it undergoes core collapse. About half of the core can be extracted safely, amounting to 23.3 million gigatons, or about 15 million years of Earth production at current rates.

Disassembly

At some point, even deep mining techniques will become impractical. Many bodies have hot interiors or liquid layers, and even without those problems, pressures would require tunnel or pipe walls too thick to be worth installing. The only approach to reaching these very deep resources is to remove all the overlying material first, which amounts to disassembling the body or a large portion of it.

The energy holding a large object together due to its gravity is called the **Gravitational Binding Energy** - . For a uniform sphere it is found by the formula

$$U = \frac{3GM^2}{5r}$$

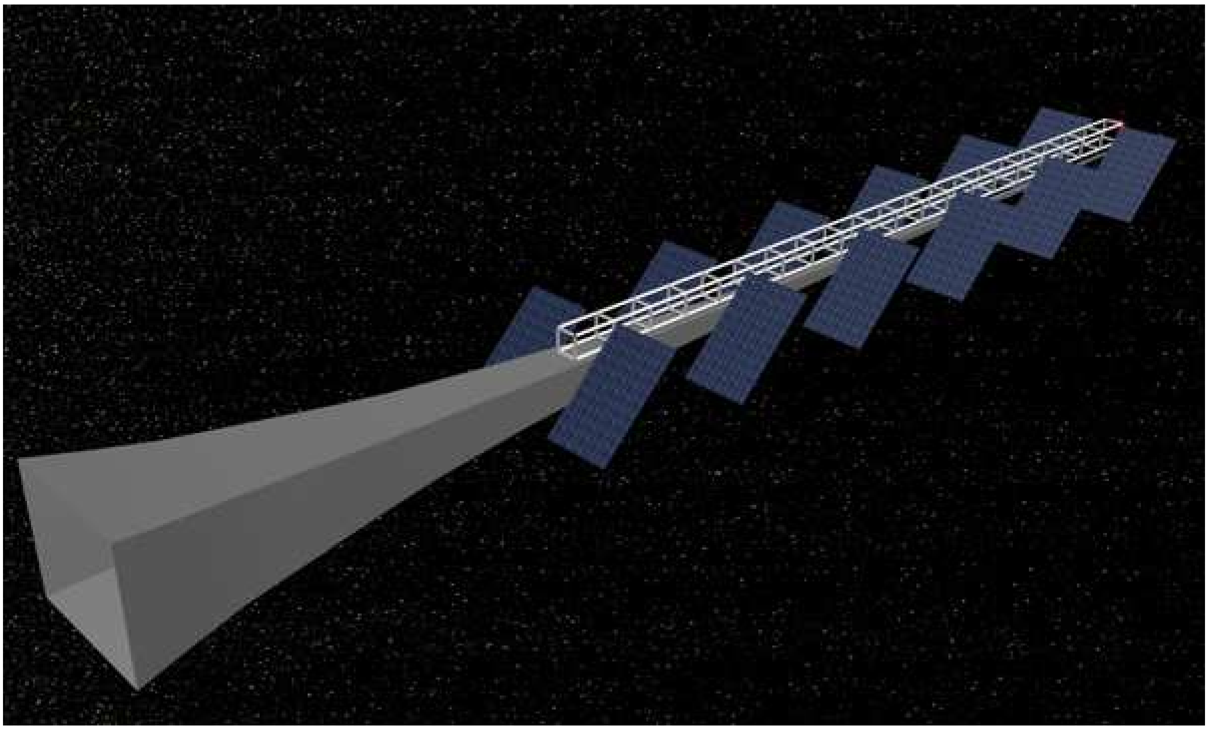
where G is the gravitational constant, M is the mass of the sphere, and r is its radius. Thus to extract all the matter from the object requires at least this amount of energy input at 100% efficiency. This assumes you don't do any processing, just physically remove all the material. As an example, for the Earth, allowing for the actual distribution of mass and density by depth, U = $2.487 \cdot 10^{32}$ J.

Dis-Assembly Time - This is a characteristic time found by dividing the binding energy U, by the solar flux falling on the object. There are other energy sources which could be used, but sunlight is generally available in space and so gives a natural value for the time to dismantle the object using the solar energy falling on it at 100% efficiency. From the characteristic value, you can make estimates of the actual time for a large mining project by multiplying by the fraction of the object you intend to extract, and dividing by the solar conversion efficiency. The solar flux for the Earth averages 1360 W/m^2, and the radius is 6378000 meters, thus the total flux is $1.74 \cdot 10^{17}$ J/s. This gives a characteristic dis-assembly time of 45.3 million years.

Mining Atmospheres

For bodies with appreciable surface pressure and a distinct solid or liquid surface, bulk mining of the atmosphere is relatively straightforward. The main requirement is a pump and a storage tank. If a particular component needs to be separated out it becomes more complex, needing liquefaction or freezing to separate the different compounds. For a low pressure atmosphere, methods like selective ionization or mass spectrometry may work to separate particular components. For bodies without a distinct surface, or in cases where you don't want to land, the orbital scoop mining method can be used.

Scoop Mining



Scoop mining ship concept

This method involves skimming the upper atmosphere of an object to collect gas, then ejecting part of the collected mass as propellant to make up for drag. The altitude to collect gas is selected based on drag and heating levels for the scoop, and available thrust. The scoop is shaped as an inverse nozzle, converting a high velocity, low pressure stream into a low velocity (relative to the scoop) high pressure gas. The collected gas is pumped into a storage tank. Since collection velocity for objects with atmospheres is typically higher than a chemical rocket exhaust velocity, the portion ejected to make up for drag will need to be at electric thruster velocities (30-50 km/s). To maintain orbit, the average thrust must equal the average drag. The ratio of scoop velocity / exhaust velocity determines what fraction of the collected gas needs to be used for thrust. The concept will work in theory for any body where the exhaust velocity is sufficiently higher than the orbit velocity.

Intermittent Operation - Take as an example mining the Earth's atmosphere from orbit. At an altitude of 150 km, the density is 3 x 10^{-9} kg/m^3. A vehicle traveling at 7800 m/s relative to the equator, which is 450 m/s above circular orbit velocity at that altitude, will encounter 23.4 micrograms of air per square meter per second, and see a drag of 0.18 N/m^2. The stagnation pressure if the scoop fully brings the airflow to a stop is 1350 Pa (0.2 psi). A vacuum pump pulls this gas into a storage tank at higher pressure. Assume the scoop system has a mass of 100 kg/m^2. Therefore it will lose 400 m/s of velocity in 220,000 seconds (2.6 days). The orbit is elliptical, so the collection will only be at full pressure at the lowest point, and actually take about 4 times longer. The scoop will have collected 5.2 kg of air per square meter over this time. Before too much drag happens, the scoop raises its perigee by 50 km using an electric thruster to stop the collection. This requires 15 m/s of velocity. Raising apogee by another 400 m/s, and then dropping perigee by 50 km to start the collection cycle gives a total velocity of 465 m/s required. At an exhaust velocity of 50 km/s, performing the velocity changes will consume 0.98 kg of propellant, which we assume to be some of the collected air. This leaves us with a net of 4.2 kg collected.

Continuous Operation - For a second example, assume a VASIMR type thruster which uses 200 kW power to produce 5.7 N thrust. The solar arrays need to be about 13 x 42 m in size, broken into 4 smaller arrays of 13x10.5 m which each follow the Sun. They are arranged lengthwise along the orbit direction, with the thruster in the back. Assume the thruster works no more than 30% of the time while mining, to allow margin to eventually raise orbit. Therefore the average drag can be 1.7 N. Assume a 13x10.5 m scoop at the front, which matches the solar array dimensions, thus a 136.5 square meter area. The atmospheric density must then be 2.3 x 10^{-10} kg/m^3, which occurs at 200 km altitude. The scoop will collect 0.23 grams/sec, and the thruster will consume 0.034 grams/sec, leaving a net of 0.196 g/s (16.9 kg/day). Eventually whatever storage tanks you use are full, and the mining ship increases thrust and climbs to an orbit where it is not seeing significant drag. If the tanks hold 5 tons of air, the mining ship can collect it in 300 days. The question is then what the mass of the mining ship, to determine a mass return ratio.

Design Concept - The illustration shows a general concept for the scoop miner. The scoop is shaped as a hyperbolic cone and functions similarly to a turbo-molecular pump by bouncing incoming molecules off its surface. Once the density is sufficiently high for the air to act as a fluid instead of individual atoms, a stagnation region will form at the narrow end, which a conventional vacuum pump can collect from. Solar arrays are mounted behind the scoop so as to not increase drag, and are pivoted to follow the Sun as the vehicle orbits. At the rear are a storage tank, and the electric thruster powered by the arrays. As a safety measure, the vehicle should carry conventional thrusters and fuel to raise orbit if the main electric thruster fails.

Trolling for Air - Trolling is meant in the fishing sense, and not the annoying Internet person sense. In this version, the electric thruster and main solar arrays are at a higher altitude, and so not limited in size and power by drag limits. The scoop, pumps, smaller solar array to power the pumps, and storage tank are lowered on cables to an optimum altitude to collect air. If the same size as the continuous operation version above has an empty mass of 1 ton, then it needs to be at least 1 km below the center of mass for the gravity gradient (tide, or differential gravity) to be larger than the drag force. It will then hang at a trailing angle balancing gravity and drag. To have enough difference in air pressure to be worth lowering the scoop, the cable will be much longer than 1 km. When the tank is full, the cables are reeled in and the tank is unloaded or swapped. This saves using propellant for climbing up and down from the scoop altitude, and allows more powerful thrusters. Whether it is an overall advantage requires a more detailed study

Terrestrial air scoops were studied in the early 1960s by R.H. Reichel at Boeing. This work looked at the operation of a scoop in continuum flow (110 km altitude) and in a free molecule flow environment (160 km altitude). Also examined were the power requirements for liquefaction as well as radiator area requirements. The famed Michael Minovitch also studied terrestrial air scoops in detail. His 1988 patent for a self-refueling space propulsion system deals with a scoop that would operate in working fluid environments and operating between 50 and 100 km altitude.

References:

- Reichel, R. H.; Smith, T. L.; Hanford, D. R. "Potentialities of Air-Scooping Electrical Space Propulsion Systems", presented at the ARS Electric Propulsion Conference, Berkeley, California, March 14-16, 1962

- Minovitch, M. A. Self-refueling Space Propulsion System And Operating Method. Patent 4,754,601. July 5 1988. Print.

Mining Liquids

A few bodies, the Earth being a notable one, have surface liquids, and others are known or suspected of having subsurface liquids. Collecting surface liquids in bulk is again a straightforward matter, needing a pump and storage tank. If the liquid is a mixture of compounds and a particular component is desired, then physical or chemical processes need to be applied to separate the desired component.

Drilling into a body with a submerged ocean, like Europa, might be hard, but seismic measurements should not be. Place one or more seismic detectors on Europa's surface, and then smack it really hard, like with the upper stage that launched the probe, and look at the vibrations you get. That's how we know about the interior of the Earth and prospect for underground resources. Boundaries like ice/water tend to reflect vibration waves (ie sound).

There will likely be natural vibrations from Europa flexing and moving (Europaquakes), but you can't be sure of them, so better to have your own source of vibration from a high speed impact.

Mining Gas Giants

Gas Giants have no solid surface. At a particular depth their atmospheres gradually become supercritical fluids, rather than a distinct layer, so different mining techniques are needed, especially if you want to access deeper materials. Scoop mining has been described above, for skimming the upper atmosphere. Buoyant mining equipment is possible but difficult. Gas giant atmospheres typically are mostly Hydrogen and Helium, so it is difficult to design equipment lighter on average than the lightest two elements. Reaching orbit is also difficult due to the thickness of the atmosphere and depth of the gravity well. To access deeper materials, large scale and rather drastic methods would be needed such as spin-up, boil-off, and disruption. The last is not recommended for inhabited stellar systems.

Spin-Up

If scoop mining is insufficient in volume, and buoyant mining is too difficult, this method increases the typically fast rotation rate of gas giants until the equator is moving closer to orbital velocity. This makes removal of material from the equator to orbit easier. There are a number of techniques for increasing the rotation rate:

- **Aerobraking momentum transfer:** - A vehicle accelerates while in orbit around the gas giant, then deposits excess velocity by braking in the upper atmosphere, adding momentum to the planet.
- **Kinetic deposition** - Very high speed objects are directed at the equatorial region of the gas giant, such that the average rotation rate increases. Depending on impact velocity the impactor may be absorbed by the planet, or material may be kicked out as ejecta.
- **Tidal coupling** - Intentionally place one or more sub-synchronous satellites in low orbits to raise tides which will accelerate the gas giant's rotation. This method is very slow
- **Magnetic coupling** - If the gas giant has a magnetic field, react against that field to spin up the core, and eventually the rest of the planet.
- **Reaction motor** - A large high velocity fusion engine, powered by Hydrogen from the gas giant's atmosphere, is mounted in the upper atmosphere, and accelerates itself against atmospheric drag, which ends up adding to the planet's rotation. For this to function, the exhaust plume needs to be able to leave the planet.

Boil-Off

This method involves reversing the way the planet formed in the first place. Planets form by collapse of a gas cloud as it radiates away energy. The largest Solar System gas giant, Jupiter, is apparently still radiating away excess heat today, after 5 billion years. If excess solar energy is directed at a gas giant, it will heat up and reverse this process. Assuming the outer layers heat up faster than the interior, this would puff up the atmosphere into regions where orbital and escape velocities are lower.

Mechanical Disruption

This is a brute-force method. One approach involves directing a large body at high speed at the planet. The other approach is to collect hydrogen, deuterium, or helium-3 and use them to make a very large thermonuclear device. The end result to to throw large amounts of material into planetary orbit or entirely disrupt the planet. It will be difficult to get material to only go into orbit and not escape completely, which is why this method is not recommended for systems that are already inhabited.

Mining Particulates

Stellar systems and interstellar regions contain small objects, particles, and gases in low density distributions. These include:

- Asteroid rocks smaller than 6 meters diameter and mass below 200 tons, for which collecting is a more apt term than mining.
- Smaller particles such as those which create the **Zodiacal Light**, coming from asteroid and comet debris, which range down to 10 micrometers in size.
- A flux of particles called a **stellar wind** emitted from stars.
- The interstellar medium, consisting of very low density gas and dust grains
- Interstellar gas clouds and nebulae of higher density than the general interstellar medium, but still low density by ordinary standards.

The challenge for mining these types of materials is their low density, forcing large collection volumes in order to collect significant amounts.

Mining Stars

If a civilization's material needs are large enough, it may consider mining the stars themselves, where most of the mass of a stellar system is located. The temperature of stars is an obvious difficulty in doing this. The following methods are speculative at this time:

Artificial Red Giant

Class M stars have surface temperatures below the melting point of the most refractory compounds, and so might be mined directly. If you surround a star with reflectors such that most of the energy is trapped, the outer layers of the star will heat up and expand, creating an artificial red giant. When sufficiently expanded, you can skim off material and take it elsewhere to be used

Plasma Jets

Artificial magnetic fields might be used to create flares or jets from the star. When sufficiently far away the material is collected and used.

Stardiver

A sufficiently large planet is placed in a highly elliptical orbit which grazes a cool star. The planet will collect an atmosphere by gravity during the grazing.In the outer part of its orbit it will cool down and can be mined, and the process repeats. The required size of the planet is such that it does not lose mass from evaporation during the close pass.

Energy

There are numerous energy sources in the environment near stars. They are listed below in approximate order of available power and ease of extraction. There are fewer significant sources in the dark regions between stars, but we will consider what is available.

Sun/Star Energy

Photovoltaic Energy

Photovoltaic conversion has been by far the most common way to extract energy in space. This is due to simplicity, scalability, mass, availability, and durability. Photovoltaic cells generate electricity directly without additional conversions or devices, They come in typical units of a few watts each, which can be scaled by simply using more of them. They are fairly low mass per power produced, and do not consume fuel. Except in the shadow of large bodies they can produce power 100% of the time, and have operating life measured in decades except in high radiation environments.

Cells for use in space operate on the same principles as ones made for use on Earth. The operating conditions are different, so they are somewhat modified. Sunlight is not filtered by the atmosphere, so they work with a different spectrum. Temperature ranges, vacuum conditions, and higher UV and other types of radiation all have to be accounted for. If higher power levels are needed, and launched from Earth, panels of multiple cells need to be folded for launch, and then deployed. Spacecraft pointing needs are typically different than the Sun's direction, so the panels have to be articulated so they can follow the Sun. If used in the shadow of a large body, some storage in the form of batteries or other devices is needed, or power must shut off temporarily.

Thermal Energy

Wind Energy

Fission Energy

Fusion Energy

Orbital Energy

The potential and kinetic energy of bodies in orbit is a potential source of energy.

Section 3.5 - Processing and Production

The **Resource Extraction** section discussed how to extract raw materials and energy from the available resources. This section covers how to convert those into finished inventory and components which can be assembled into completed systems and projects.

Existing industry extracts materials and energy, produces bulk supplies and parts from them, and assembles them into finished items entirely on Earth. The methods, however, are independent of location. We just happen to have evolved on Earth, and so that is where industry first was set up. In principle, the same methods used on Earth can also be used in space or on other planets and objects. In practice, we specialize industry on Earth for location where it functions best, and the same will be true for locations beyond Earth. There are some additional methods which only are suited to locations beyond Earth, due to special conditions like available energy, lack of gravity, or vacuum. A 1993 survey of methods and uses specifically for space, **Resources of Near Earth Space** represented the state of the art at that time.Progress has been made since then.

This section will list the full range of available methods, but thought needs to go into selecting which to use in a given location and circumstance for a space project. Some Earth-based methods assume gravity or air pressure in their operation. Artificial gravity or atmosphere can be provided in space locations where it is not naturally present, but there is overhead and complexity in doing so. Thus in designing a space production facility, you should consider an alternate process that does not require special conditions, rather than automatically adding them to a traditional process that needs it.

The scope of all known production methods is too wide to fit in the confines of a single book. Indeed, the entire fields of Industrial and Chemical Engineering are devoted to this topic. We will give a summary of the available methods, with pointers to more detailed information. A typical industrial process uses multiple steps in series or in a more complex flow of operations with branches or loops, under some method of production control. Here we list individual steps, which are called **Unit Operations** in process engineering. The range of possible complete processes by combining these steps is quite large. The task of the system designer is then to select the proper set of steps and complete processes for the task at hand.

Production Control

Processes do not operate themselves. Under the heading of **Production Control** fall those elements that plan the production, send commands on what to produce to the system elements that actually perform the work, and monitor the status and outputs of the operations. Control can be either manual or automated, and either local or remote depending on circumstances and design. Processes can be repetitive or continuous, producing the same output at some average rate, or singular, where a different part is produced each time.

Computer Design

Process Measurement and Control

Automation and Robotics

Handling and Storage

Materials and parts need to be moved from process to process, or stored between processing operations or when complete.

Parts Transport

Parts Storage

Bulk Transfer

Bulk Storage

Factory Environment

Materials Processing

Materials processing converts raw materials from their state as delivered from mining to finished bulk materials such as water or oxygen, or ready stock such as bars, rods, or sheet.

Mechanical Processing

Mechanical processing changes the physical state but not the chemical composition of the materials

Crushing - This is breaking down a material into smaller pieces by applying pressure. **Milling** refers to making finer powders. In addition to standard methods of crushing, kinetic impact can be used as a process in space, one that Nature has applied extensively.

Sorting - This is the sorting of material by size or type using gravity, acceleration, vibration, electrostatic forces, or magnetic fields. Besides physical wire mesh or perforated plates to sort by size, methods like electrostatic can sort materials by charge-to-mass ratio.

Mixing -

Thermal Processing

Thermal Processes

Evaporation

Condensation

Heat treating

Crystallization

Drying

Heating and Cooling

Thermal Sources

Microwave Heating

Refining and Separation

Filtration - Separation of solids from liquids

Distillation

Chemical Processing

Ore reduction

Alloying

Reaction

Synthesis and polymerization

Organics Processing

Agriculture and Food Growth

Organics Conversion and Storage

Parts Production

Forming and Molding

Molding

Blowing

Casting

Rolling

Forging

Subtractive Fabrication

Mechanical Machining

Shearing

Sawing

Drilling

Milling

Abrasives

Electrical Machining

Electric Discharge Machining

Electro and photo chemical machining

Plasma arc machining

Beam and Jet Machining

E-beam & ion milling

Laser cutting

Abrasive water jet

Additive Fabrication

Extrusion

Vapor deposition

Powder forming and sintering

One method similar to sintering is to spray coat reinforcing fibers with droplets of molten metal to form a reinforced tape. Then the tape is applied in layers to build up the shape you want. Each layer is heated to just below the melting point then pressed to the previous layer to bond it.For making shapes like cylinders this can be a continuous winding process.

Gluing

Welding

Brazing

Soldering

Fiber spinning

Weaving

Sewing

Coating

Painting

Coating

Printing

Plating

Dyes

Electronics Fabrication

There are three main levels of electronics fabrication. These are making the individual components, connecting those into circuits, usually in board form, and assembling them into finished units.

Component Fabrication

Circuit Fabrication

Unit Fabrication

Section 3.6 - Assembly and Construction

Assembly refers to putting together parts or component systems to make a functional complete device. Construction refers to permanent installations not intended to move often or ever and consisting of many devices and systems. These tasks typically consist of many simple steps in series.

Assembly Methods

Assembly methods can be organized by the type of operator and the type of connections are made between elements

Operator Types

Assembly methods can be categorized by the type of operator that performs the assembly:

- **Manual Assembly** - This is performed by humans using manual dexterity and strength. It can be assisted by tools and fixtures for leverage, speed, and positioning.
- **Robotic Assembly** - This is performed by robots either under human remote control or computer control.
- **Automatic Assembly** - This is where the devices to perform the assembly task are built into the components.

Mechanical Connections

Electrical Connections

Fluid Connections

Construction Methods

Construction can be organized by analogy to Earth construction, with suitable modifications

Construction Planning

Logistics

Delivering construction equipment and supplies to the construction site from other locations.

Orbital Tugs

When launching multiple components from Earth that need to be collected in one place for assembly, if the orbits are similar, then a single vehicle dedicated to the job is more efficient than including a propulsion and navigation system on each component payload [1]. We call that vehicle an **Orbital Tug**. Besides propulsion, navigation, and means to grab the payloads, it needs the ability to periodically refuel. This can be done with occasional payloads being fuel tanks, which the Tug attaches to itself as needed. In order not to be left stranded, it should have at least two tank locations, and install a fresh tank while the second one still has some fuel left. If you have more than one Tug, they should be designed to grab each other for heavier cargos needing more total propulsion, or in case one stops working and needs to be returned for repair

Materials Handling

Positioning items within the construction site to the point of use.

Docking

Mechanical Docking Systems

Deployment

Unpacking a component shipped in a compact state to an operating configuration.

Site Work

Where you are doing construction on or near a natural body, and you need to modify the construction area by road-building, excavation, and other methods.

Cast and Block Structures

These are not pressure tight structures, but installed by methods like casting concrete or stacking blocks

Fabricated Structures

These are from metal, wood, plastic, or other materials delivered as components to be assembled on site

Outfitting Methods

Outfitting involves installing and activating items within a larger construction element once it is in place.

Sealing and Protection

Utilities Installation

Equipment Installation

Habitation Setup

Activation

References:

1. Gralla and De Weck **Strategies for On-Orbit Assembly of Modular Spacecraft** (http://strategic.mit.edu/docs/2_20_JBIS_60_6_219_Tugs.pdf), JBIS, vol 60, p 219, 2007.

Section 3.7 - Verification and Test

Verification is proving that a system element or the system as a whole meets the requirements imposed on it. **Test** is one method to gather data for such proof. Other methods include analysis and inspection. Since test is an integral part of most verification processes, they are normally grouped together. Verification and test is normally applied at multiple levels of a design, starting at the component level and working up to the complete system. This is to discover problems at the lowest level possible, where it is easier to identify the cause. If lower level problems are resolved, any problems that occur at the next level of integration are likely to be from the integration of multiple components itself.

Section 3.8 - Operation and Maintenance

Prior space projects have relied on a **Mission** concept of operations. That is you plan a mission ahead of time, with a detailed timeline of all the steps you expect to perform. Then the crew and ground controllers execute the plan. If deviations happen, corrective action is taken to return to the plan as much as possible. This concept works for a single mission in isolation with a well defined goal. Future projects will have multiple activities in multiple locations, with larger numbers of people involved and interactions between activities. So planning every step ahead of time in a linear fashion becomes difficult if not impossible. For cost reasons, you also want to reduce the large amount of planning and training that currently goes on relative to the actual doing.

Operations Concepts

For future projects, rather than planning everything in advance, one approach is to develop building blocks of tasks which can then b assembled as needed to reach a goal. As new feedback comes in from other tasks, the building blocks can be adjusted or replaced as needed. Changing plans may not be as efficient as carrying out a fully detailed plan made in advance, but the effort to make those advance plans may be larger than the efficiency loss. The ability to change or replace tasks, rather than a monolithic preplanned mission, also allows continuing operations to adapt and improve over time.

Operations Tasks

Debris Removal

This task is the removal of human-made orbital debris, and, if possible, natural hazardous objects. Over time, debris will collide with itself and decay from drag. In the long term it will clean itself up, but in the short term the collisions will generate more hazardous size pieces, which is an unsatisfactory situation. Besides clean up of past debris, this task takes care of cleaning up future defunct, lost, or destroyed hardware. There are a number of methods which have been proposed to do this, none of which have been demonstrated yet:

- A "plasma-puffer", promoted by Daniel Gregory. That would use an arc discharge from a high altitude balloon to puff air upwards in the path of the junk, making it de-orbit faster
- Ballistic puffer, which launches a projectile from a very large gun, that releases air or some other material in the path of the debris, having the same effect. The projectile is sub-orbital, so it does not add to the space junk problem.
- Lasers - zap the front of the debris, blasting off some of it, and slowing it down from the reaction.
- Drag devices - attach a large sail or conducting wire to the space junk and make it slow down faster
- Orbital tug - goes from orbit to orbit collecting junk. This requires high efficiency electric thrusters or the fuel used gets absurdly high.

Maintenance Concepts

Maintenance Tasks

Section 3.9 - Recycling Methods

Most locations off the Earth do not have automatic recycling of waste products, and a space project will have a finite amount of materials that have been processed beyond the raw state. We define **Recycling** as the application of energy and processes to convert materials from a non-useful state back to a useful state. It shares some technology with production from raw materials, except using materials that have already been processed previously and are not in their raw state. In fact some recycling methods will involve feeding waste products back into a production process.

Waste Recycling

A goal for recycling is to achieve **Closure**, where all waste products are converted back to useful states and the only external input is energy. We expect this to be a theoretical limit similar to conversion efficiency in electrical transformers and motors. Those devices can reach 98% efficiency but not 100%. Similarly, practical recycling is expected to reach a high percentage, but not 100%. To the extent your percent closure CL% approaches 100, the amount of new raw materials required is reduced to 100-CL% of an open system with no recycling. So a high percent closure can have a dramatic effect on the need for new raw material processing or replacement items brought from elsewhere.

Some items by their nature are not amenable to recycling. A prominent example is reaction mass expelled outside a gravity well by a propulsion method. You are deliberately throwing that mass away in order to get thrust, and outside a gravity well there is no practical way to recover it to use it again. Within a gravity well, such as launching from Earth using a chemical rocket, all the reaction mass is sub-orbital - ranging from about -1/2 orbital velocity to +1/2 orbital velocity depending on the vehicle velocity. Therefore it all ends up back on Earth and can be used again. While some reaction mass cannot be recovered, you can deliberately choose high efficiency methods that lose less mass this way, and use methods such as gravity assist that do not lose propellant mass.

Closed Loop Life Support

Humans are biological organisms with certain requirements to continue living, including clean air, water, and sufficient food. They also produce waste products from the human standpoint, such as CO_2, urine, and feces. On Earth those wastes from the human standpoint are necessary inputs to plants and other organisms, and solar powered evaporation produces clean water in the form of rain. The natural cycles form closed loops. For space projects we have to replicate the function of those closed loops with artificial systems, or supply air, water, and food from external sources. The more time and the higher the number of humans the greater the outside supply mass becomes, and so the more desirable a closed loop system becomes.

Besides inputs such as air, food, and water, humans additionally need controlled lighting, temperature, pressure, radiation levels, and acceleration. Prepared food and feces are complex from a chemical standpoint, there are secondary volatiles and shed skin produced by humans, and body and clothing cleaning are desirable. Therefore a full life support system for humans is complex. We apply the Systems Engineering approach of dividing it into simpler sub-systems that each perform part of the total functions required, and then optimize the full system as a whole. In addition to the direct functions that provide life support, such as growing plants producing food, there are also indirect functions caused by meeting the requirements of the plants, such as water, illumination, gravity, and $C0_2$ concentration. The methods listed below include meeting both direct and indirect functions.

Artificial Gravity

Humans and plants are evolved in a 1 g environment. Without gravity, human bones deteriorate and some plants may not grow properly. With gravity, water circulation and dust settling operate by familiar methods. If local gravity from a large satellite or planet is not sufficient, artificial acceleration can be generated by rotation.

Illumination

Photosynthetic plants need sufficient illumination of the right wavelengths to grow, and humans need sufficient light to see and navigate. Having both evolved on Earth with the Sun as the main source of illumination, both use the wavelength band that the Sun emits the most power at, known as **Visible Light**. There are slight differences in intensity, day cycle, and wavelengths for various plants, and for what humans need to see properly. Natural sunlight is easily obtained at many locations in space, so it certainly should be considered if available. Artificial lighting is available for circumstances where natural sunlight is not sufficient, but that usually requires a power source to operate. For a large area of plants that amounts to a lot of power, so natural sources are usually preferred, even if it needs to be concentrated.

Soil

Plants on Earth often grow in soil, which is broken down rock plus microorganisms, water, nutrients, and organic matter. Hydroponics dispenses with the inert rock component, which can save mass for space projects. If radiation shielding is required, though, the inert rock component can serve a dual function. While fully developed soils are only known from Earth, there are many sources in space of small rock particles, water, and carbon which make up a large part of soil mass.

Food

Human food requirements can be measured in terms of total energy, commonly measured in **Food Calories** which are equivalent to 4184 Joules of available energy. The requirements are also measured by a wide variety of specific nutrients in specific amounts. It may not be efficient to supply all the low mass nutrients in a given situation, while supplying the main ones by mass. Whatever nutrients are not obtained from the closed loop system would have to be supplemented from outside, or simply done without for short missions. Human requirements can be translated to specific growing areas per person using known data on agricultural productivity. This data can be modified by designs of space systems. For example, if a plant needs 12 hours of sunlight per day, you can grow twice as much in the same area by using trays that are swapped every 12 hours, and storing the other tray under the illuminated one.

Part 4: Projects and Programs

In previous parts of this book we have discussed individual systems which carry out purposeful functions. In Part 4 we will consider more complex projects and programs, which involve multiple systems. Multiple systems can exist and interact with each other at one time. They can also grow and evolve as a group over an extended time, with new systems replacing older ones. When multiple systems are directed at one or more common goals, we refer to them as a **Program**, such as the civil space program of the US.

An optimized program often results in a design with multiple systems. Reasons for this are described below. For design and management purposes, a large and complex program can be organized into a multi-level structure from the program as a whole, to individually designed systems. We then need ways to describe the parts of the program. Some common descriptions include:

- **Project** - A set of systems that exist together at one time under one management, and have a set common goal.
- **Segment** - A set of systems within a program with diverse management and shared goals.
- **Phase** - Part of a program which exists for an extended time, with significant growth, evolution, and replacement of older systems with newer ones.

Projects, segments, phases, and other pieces are then assembled to form the larger program. The names and structure are somewhat arbitrary, and are selected according to the needs of a particular program. It is important that the selected structure cover all the work needed for the program, and that all the people working on it have a shared understanding of how the structure and its parts fit together. Despite the complexity or length of a large program, the principles of Systems Engineering can still be applied to optimize the overall design. When the program's life cycle is long relative to changes in technology, society, and the environment, the Systems Engineering tasks may be repeated, or performed continuously to get the most benefit.

Reasons for Multiple Systems

In general, the more complex, or more extended in location, volume, traffic, or time a project is, the more likely it will result in multiple systems. Different environments and available resources will drive different local solutions. Changes in technology and economics over time will also guide changes in design. In any large or long-lived program, a multiple system approach should at least be considered to see if it gives a better result. Some specific reasons include:

- **Non-Linearity** - A given transportation or engineering method often has a non-linear term in it's equations - at least quadratic if not an exponential. For example drag goes as the square of velocity, and the rocket equation increases mass ratio and propellant as an exponential function of velocity. It is often is more efficient to break up the total job into components because the sum of smaller non-linear terms is less than a single larger term with an applied exponent.

- **Complex Needs** - People have complex needs, and projects we wish to accomplish typically have multiple goals. This drives design solutions which use multiple materials, devices, energy sources, etc. Therefore no single system or technical solution is likely to best satisfy all the desired ends.

- **Economics** - A single large "all or nothing" type monolithic system, which requires a large up-front investment, often turns out to be wasteful. Aside from the non-linear effects noted above, we cannot predict future technology developments, and large projects often have long development times. If you build in smaller steps, you have the opportunity to change direction if new improvements come along, or retrofit a change to just the part that needs it. Finally, with an incremental project, you can get some use from it earlier. This can produce a higher economic rate of return.

The choice between single or multiple systems should not be done arbitrarily ahead of time. The decision should, if possible, be made by analysis of the alternatives and choosing the best one. Sometimes the right choice may not be possible for non-engineering reasons. For example, the civilian space program in the US is operated as a whole by NASA, and funded by congressional appropriation. A single agency funded from a single source can make it more difficult to divide or set up separate systems, even if it makes sense from an engineering standpoint. The organizational structure, budget review process, and national politics tend to favor single and conservative solutions. Conversely, the historical division of US government-funded space activities between NASA, and the Defense, Commerce, and Energy departments can make it more difficult to combine projects and programs, even if they would be more economical.

Organization of Part 4

A major purpose of Part 4 is to show how the various engineering processes and technologies from earlier parts of the book can be put to use. We therefore present an extended example of a long-term and complex program that involves multiple terrestrial and space systems. Unlike printed paper books, the example in this book does not have to remain static, but rather can be developed over time. The completed portions will show by example how calculations and decisions are made. There are also opportunities for individuals or teams to develop the ideas further, and practice their design skills. Their work can be recorded as design studies in Part 5, or as separate documents to be linked. The best ideas can be incorporated back into the main discussion of the book. It is hoped this approach is a useful teaching method, a way for readers to gain real design experience, and a way to make real progress in in future programs.

Status - The book as a whole is about a 65% complete first draft as of Sep 2017. Part 4 is still incomplete, but is undergoing major revision as of late

2017.

Our Program Example

Our chosen example has the overall goal of upgrading and expanding civilization on Earth to more difficult environments, then to the more distant and difficult regions of the Solar System, and eventually beyond. We will leave choosing a catchy name for this multi-region effort to others, and just refer to it as "our program". The program's component parts are not intended to be under a single centralized control. This is both to illustrate the interactions between distributed program elements, and because it is impractical to run such a large program as a single entity in the real world. Each part of the program then needs its own reasons to proceed, whether they are economic have other motivations. The program's description and supporting concepts help provide these reasons, and a structure to inform and coordinate independent efforts. The various parts would then interact with each other, and with the outside world. The interacting parts, and the program as a whole, are not static, but evolve over time.

We chose this example for several reasons:

- As a long duration and wide ranging program it allows us to demonstrate design considerations and methods for many types of systems and subsystems.
- In later design studies and separate documents we can teach by example, showing in some detail how the calculations and decisions are made to arrive at a design.
- This is intended as a realistic program concept, incorporating the best current technology and ideas. If good enough, the proposed elements may actually be adopted and built.
- An open-ended program using new concepts and technologies allows readers to do useful original work. At the same time they can gain individual design skills and experience, and practice working in diverse teams.

The program concept is partly based on work by **Dani Eder**, the original author of the **Canonical List** from which this Wikibook originated. It also includes many new ideas not yet pursued by government or commercial programs. In its present state it is not complete or claimed to be the best possible program concept. Rather it is intended as a starting point using recent ideas for space development, from which an optimal design can evolve by the contributions of many other people. We intend to use the engineering methods described in Part 1 in further developing the proposed program.

Part 4 Contents

A large and complex program naturally can't be described in a single page, or even a single book. The sections and pages in Part 4 provide an introduction to the program at the concept exploration level, in roughly chronological order, according to program phases. Design studies in Part 5, and related information elsewhere, will provide more details about design choices and calculations. Current sections and pages include:

- **Section 4.1: Program Overview** - which ummarizes the program concept as it has been developed so far
- **Section 4.2: Phase 0 - Research & Development** - identifies new or improved technologies and methods that are needed for the later program phases, and how to develop them. Such items must be developed before they can be used, so Phase 0 logically precedes the other phases. When needed items are identified, they are fed back to Phase 0 for planning and integration with other work.
- **Section 4.3: Phases 1 to 3 for Earth** - summarizes improvements and expansion of civilization on Earth by means of **Seed Factories**. These are systems which bootstrap from a starter set and grow to whatever size is needed using smart tools, local energy and materials, and process and design knowledge. Such systems will later be used in the space parts of the program, but their application to Earth is outside the main topic of this book. Those applications and the seed factory concept in general are discussed in a separate book.
- **Section 4.4: Phase 2B - Industrial Locations for Space** describes the subset of industry on Earth needed to access and support work in space, such as rocket factories and launch pads. Other industries already exist, so the space-related ones on Earth do not have to bootstrap from starter sets. That approach is more useful in space where very little industry currently exists.
- **Section 4.5: Phase 4A - Low Orbit Development** - Phase 4 in general covers orbital regions away from the Earth's surface. Phase 4A covers further development of the region between 160 and 2700 km altitude, beyond what is already there and existing projects.
- **Section 4.6: Phase 4B - High Orbit Development** - covers further development of the orbital region from 2700 to 1.5 million km from the Earth's surface, excluding the Moon and orbits around the Moon.
- The **Table of Contents** for this book lists several other pages in Part 4: Hypervelocity Launcher, Low-G Transport, Electric Propulsion, Orbital Mining (2 pages), Processing Factory, Spaceport Network, and Interplanetary Transfer. These remain from an earlier draft of the book, and will be merged into the appropriate new sections. Some other sections for additional phases will be added later, and all the sections will be renumbered.
- **Section 4.12: Phase 5A - Lunar Development** - Phase 5 in general covers planetary systems and nearby orbits around them. The Moon is important enough in size and location that we place it in a separate phase at the level of the other major planetary systems. The Moon and farther regions are still in the science and exploration stage, and have yet to be developed further.
- **Section 4.14: Phase 5B - Mars Development** - This section covers the development of Mars and the orbital region around it beyond the current science and exploration activity.
- The **Later Projects** page is currently a place-holder for additional pages which will be added later.

Design Studies and Related Information

- **Section 5.1: Program Conceptual Design** - This is a design study showing how the conceptual design for the Upgrade Program is developed. It is more extensive than a typical technical study final report. Final reports record the results of a study. Here we also show the logic and calculations to reach the results, so that others can learn from and improve upon them.

Section 4.1: Program Overview

This section provides an overview of our chosen program. Later sections of Part 4, design studies in Part 5, and related information located elsewhere and linked will provide more details. The program concept is incomplete as of late of 2017, and readers are invited to contribute to it and improve it. Conceptual and Preliminary Design, which are the early stages of overall design and development, have not been finished. Therefore we have not selected all the preferred choices, and multiple candidates are often presented.

Program Goals

Every program needs identifiable goals and objectives to direct the design and implementation work. **Section 5.1** was an early study to develop these goals and objectives, and an initial design concept to meet them. The most compact statement of the program's goals are:

Upgrade civilization on Earth, and progressively expand to more difficult environments, including multiple regions in space.

Program objectives supporting this goal include:

- **Improving Life on Earth** by developing better technology to make material goods and to live sustainably from local resources.
- **Expanding Material and Energy Resources** by access to currently dificult Earth and space locations.
- **Increasing Biosphere Security** by adapting to more dificult environments, including future changes to the Earth itself, and by countering undesirable changes.
- **Reducing Hazards from Space** by identifying what they are, followed by developing methods to deal with them.
- **Understanding the Earth Better** by observing our home planet, its environment in space, and other planets and environments.
- **Long Term Survival** by dispersal to multiple locations and acquisition of needed new resources.
- **Increasing Choice and Freedom** by opening unoccupied regions to habitation and use.
- **Increasing Opportunity** by access to unclaimed resources and more efficient technology.

Expected benefits from this program include:

- **Low cost access to space**, removing a current barrier to activities there.
- **Spin-off technology** from attempting dificult tasks, which can then find uses elsewhere.
- **Optimism for the future** by demonstrating we are not in a finite, closed world. An optimistic viewpoint in turn changes how people act.

These benefits of the program accrue to civilization as a whole, although the specific projects may be funded and carried out by smaller organizations.

Program Summary

Our current civilization significantly uses only 13.5% of the Earth's surface. The biosphere plus the human-built environment averages about 200 kg/m^2 if distributed evenly across this area. This is a tiny fraction of the 11.71 billion kg/m^2 of the Earth's total mass. The known major planets and smaller bodies of the Solar System amount to 447 times more mass than the Earth alone. In terms of thickness, our current civilization amounts to a 20 cm thick layer on the portion of the Earth we use. This is a very thin veneer compared to the bulk of the Earth. The layer thickness assumes the contents average the density of water (which living things approximately do) and were flattened and distributed evenly.

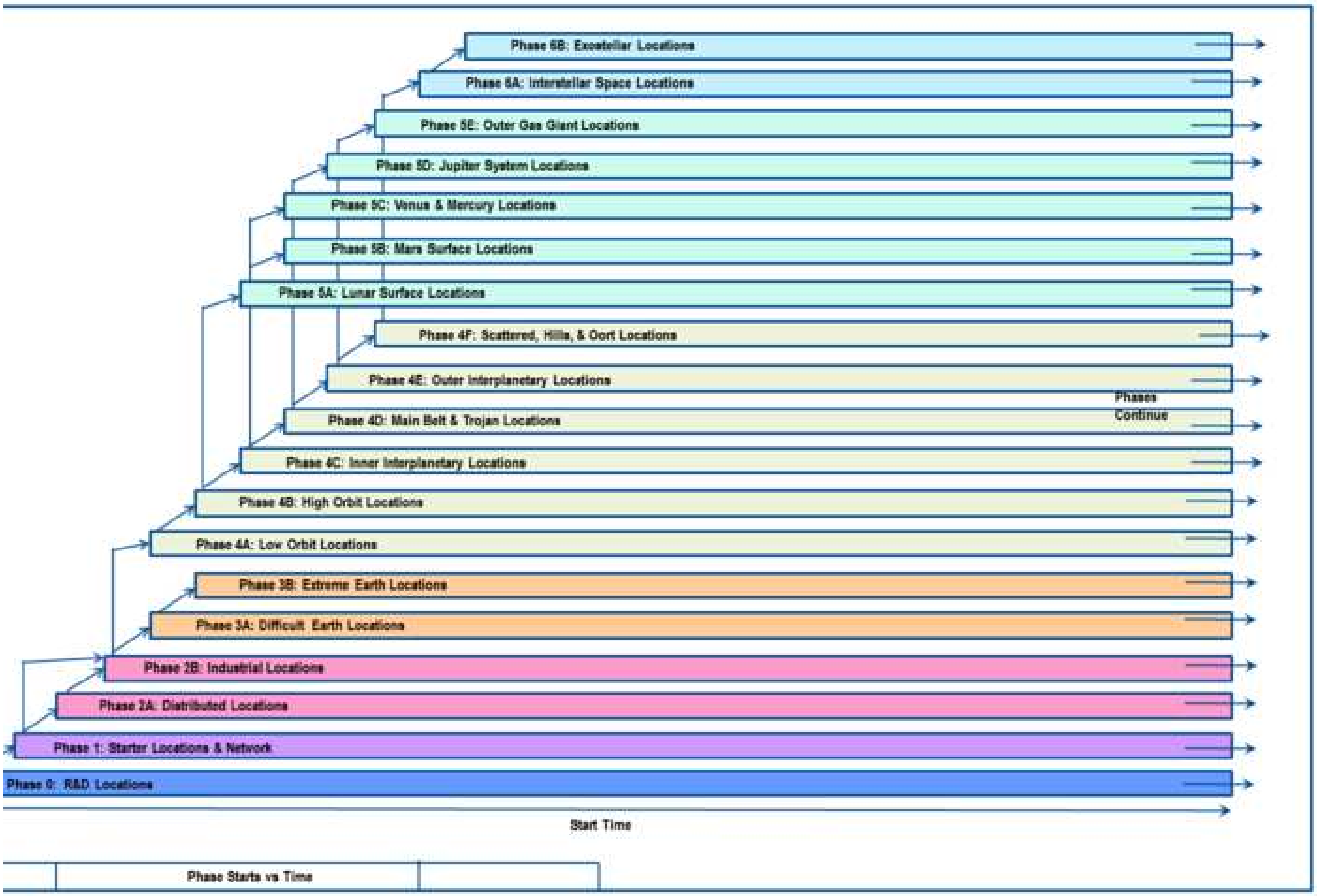
Our civilization uses about 20 TW of energy in all forms. Ignoring other energy sources, this is a small fraction of the 174,000 TW of sunlight that falls on the Earth, and a microscopic fraction of the 383 trillion TW the Sun produces. Problems like poverty and material scarcity don't stem from a lack of resources. They exist because we hardly use the vast amounts of material and energy resources that are there. Our program's approach can then be simply explained as "using more of what's already there". Of course, we want to do so responsibly, and minimize side effects to the biosphere and people's current lives.

General Approach

Historically, people first occupied the easiest environments, and used the easiest sources of materials and energy. Our program concept is to continue the "easiest first" path, but apply modern technology towards occupying and using progressively more difficult environments on Earth, and then in space. We can leverage existing production equipment and use "smart tools" (automation, software, robotics, and AI) to efficiently make more. This starts in the easiest places, which are existing locations where people already live. Starter sets of equipment are then sent to the next harder environments, where they can use local energy and materials to grow. When they have expanded enough, they can then contribute to starting in even harder locations. People accompany the starter sets when the environment is not too difficult. In the more extreme and remote locations the equipment operates automatically or by remote control until suitable space for people is built. In this way civilization can progressively upgrade itself where people already live, expand horizontally across areas we mostly don't use yet (i.e oceans, deserts, and ice caps), and vertically down into the Earth and up into space.

We expect the program would be carried out as many separate projects, each for its own reasons. For the purpose of engineering analysis, design, and optimization, though, we will consider the program as a whole. This is similar to how airplane design has to consider the airports and traffic control system they operate with, even though the airplane manufacturers, airlines, airports, and traffic control network are all operated by different organizations. By presenting the entire program, we hope to provide a hopeful vision for the future, and inspiration for people to follow selected parts according to their interests and abilities.

Our "easiest first" approach is a progressive series of upgrades and expansions to new environments, with previous regions providing the means to open up the next ones. Existing environments continue to be occupied and used as new ones are established. Therefore we organize the program into a number of phases, with staggered starting times (Figure 4.1-1) and a sequence for which phases lead to the later



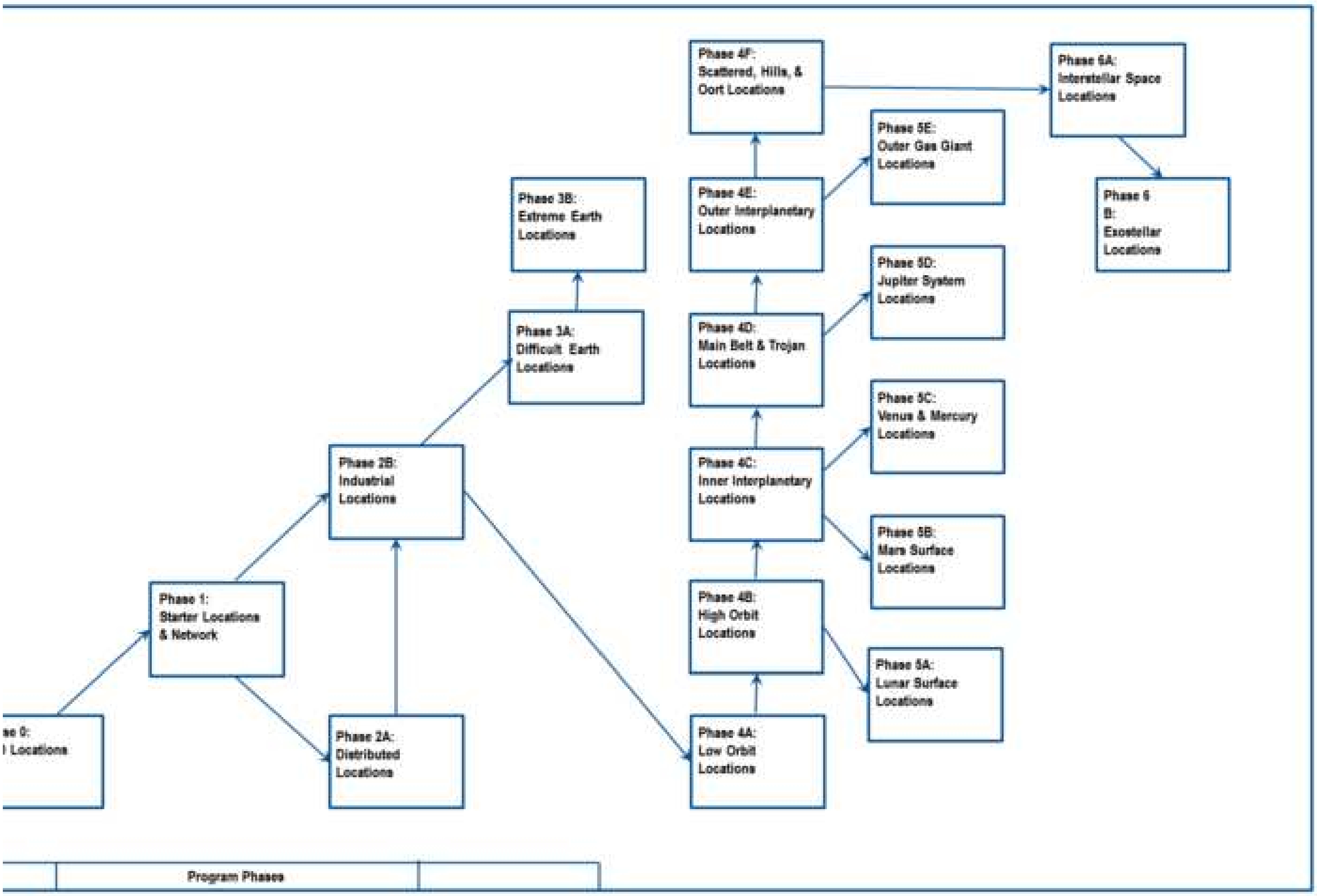
1 - Program Phases vs Tme.

ones (Figure 4.1-2). Once started, a phase continues in parallel with earlier phases, and the rest of civilization outside the program. In turn, phases are divided into smaller projects, locations, and functions, with engineered systems and subsystems which perform the functions. The parts of the program interact with each other at all levels, and with the rest of civilization, via an assortment of inputs and outputs.

The first four phases begin on Earth.These are:

- **Phase 0: Research and Development**
- **Phase 1: Starter Projects & Network**
- **Phase 2: Distributed and Industrial Development**
- **Phase 3: Other Earth Development**

The Research & Development Phase supplies necessary technology and designs to the later phases, and therefore comes first. Phases 1, 2, and 3 are distinguished by scale of operations and moderate vs difficult or extreme operating environments. Different scales and environments will lead to different specific designs, so we define phases for each. Smaller scale equipment can be used to build larger equipment, and gathered in numbers to make larger production systems, hence the sequence from starter to distributed and industrial phases. The easiest places to start are developed areas with moderate conditions, so the first three phases begin there. Phase 3 builds on that experience and therefore comes fourth, in progressively harder environments. If you wonder why four out of seven phases in a book about space systems are on Earth, keep in mind that most current space activity actually occurs on the ground. For example, six people occupy the International Space Station in orbit, but the US portion of the project alone employs about 12,000 people on Earth.



2 - Sequence of program upgrade and expansion phases.

The remaining three phases are in space, which are typically harder than the most difficult locations on Earth. This is partly due to conditions like lack of breathable air, radiation, and temperature. The energy and travel times to reach various regions in space adds to their difficulty. These phases are:

- **Phase 4: Orbital Development**
- **Phase 5: Planetary System Development**
- **Phase 6: Interstellar Development**

Orbital development begins with regions nearest the Earth, and progresses outwards by distance to higher orbits, then to ones around the Sun from inner to outermost. Planetary System regions are tied by gravity to the Moon and major planets. They include orbits around them, on their surfaces and their moons. Finally, Interstellar regions are not tied to our Sun and Solar System. The order of Phases 4, 5, and 6 is according to distance and difficulty of reaching their locations, since equipment for a given region has to come from or through the previous ones.

Most of the major phases are divided into more specific sub-phases: 2A, 2B, 3A, 3B, etc. The entire sequence of growth and expansion is intended to be self-funding once started. Self-expanding production and other new technologies should be valuable enough to justify development on their own. Using those technologies progressively on Earth and in space can fund each later expansion.

Phase 0: Research and Development (R&D)

Existing civilization has already accumulated a large amount of technical knowledge, designs, and equipment. There are some new items that are needed for the various program phases, so the **R&D phase** will do the necessary work to develop them. This includes conceptual and preliminary design, component research, detailed design, prototype fabrication, and testing. Finished items are supplied to later phases, and also to society at large when appropriate. New knowledge and technology from this phase is also contributed back to civilization as a whole. The R&D Phase is divided into multiple sub-phases, according to the later phases for which the R&D is needed. This includes for Phase 0 itself, in case unique facilities, equipment, or processes are needed to carry out the R&D work.

Seed Factories

The first major technology to be developed in Phase 0 is that of **Seed Factories**. We feel this concept is important enough, and its application to Earth distinct enough from space, to devote a separate book to the subject. Seed factories are systems which can bootstrap from a starter set of people and equipment to whatever size you need. They grow by making more equipment for themselves using existing tools and machines, local energy and material resources, and process and design knowledge. As new and different equipment is added to the collection, the growth rate can increase exponentially. Like any other factory system, seed factories are intended to make useful finished products. The portion of output devoted to growth vs finished products can be variable, and is determined by user needs. Where possible they take advantage of **Smart Tools** such as automation, software, robotics, and artificial intelligence, to leverage the work that people do. Smart tools and remote control are especially useful in dealing with difficult and extreme environments.

The current state of technology does not allow a seed factory to be fully automated or make 100% of its own parts, so it is not entirely **Self Replicating**. Instead, the goal is to reach a high and increasing level of self-production and automation. When expanding to a new environment or location, you bring a starter set, and it sets about using local materials and energy to make parts for new equipment. Whatever can't be found or made locally continues to be imported, but at a decreasing level as local capacity grows. The farther away and more difficult the environment, the greater the advantage of producing locally over bringing everything from elsewhere. So this approach becomes more useful as you reach farther into space. Even in developed locations on Earth, a starter set would be less expensive to buy than a full factory. So seed factories have economic justification for terrestrial use.

In already developed areas, the growing factories can also take local waste and scrap as inputs, besides using local raw materials. This provides environmental benefits. In both developed and new locations they take whatever inputs they get and convert them to finished materials, then finished parts and complete products. Items the factory cannot make efficiently will continue to be imported, and a small percentage of items will need rare materials not found locally. So even when it has reached full size, the factory is not likely to produce 100% of what it needs. All of the locations within the program, and the rest of civilization, will then make up a linked trading network, supplying each other with things they lack locally or are too hard to make.

Additional Technologies

Some other new or upgraded technologies will be useful for the program. One example is distributed production networks. Traditional factories brought people and equipment together in one place because it was the only way to efficiently organize the work. With modern communications and transport it is possible to coordinate work that is distributed across many places. This has been achieved for software development, but needs improvement for physical goods. Areas that need improvement include remote operations, and automatic process compilation to support an ever-changing set of equipment.

In space technology, current programs still do not fully incorporate already existing knowledge and hardware, such as electric propulsion. They don't yet incorporate the most basic in-space production methods, although some research is ongoing towards it. Many new ideas languish from lack of funds. So the R&D phase will follow a dual approach. First is to make the best use of what already exists in planning a coherent program. Second is to use some of the production capacity we develop on Earth to build and test equipment for the new and yet-untried ideas. Many of these ideas are listed in Parts 2 and 3 of this book. Since some of the space technology can only be tested fully in space, later parts of the R&D phase will involve locations in space.

Phase 1: Starter Projects & Network

Early Phase 0 R&D, Phase 1, and Phase 2 all occur in moderate environments on Earth, which are the easiest places to start. We define moderate conditions for a number of environment parameters such as temperature, water supply, and air pressure. Each parameter is considered moderate where the middle 90% of people currently live, with 5% at each extreme. If any parameter is outside the moderate range, then the whole location is no longer moderate. The moderate range is intended as "typical" or "normal" conditions to design for, such that a single design should be able to operate in any part of the environment. Moderate conditions are measured by the natural exterior environment, and not the interiors of controlled production, habitation, or transport spaces.

Phase 1 begins gaining experience with bootstrapping from starter sets by building the first operational seed factory equipment. Early equipment includes conventional tools and machines that are bought or made, plus starter set machines designed in the R&D phase. Starter locations are in already developed and populated areas which can supply whatever else is needed to get started and operate. Phase 1 machines are typically small, such as for hobby crafts and home improvement type uses. This makes them affordable for individuals or small groups. The machines can be located in homes or locally-built workshops. Besides direct use for small-scale production in this phase, the experience will be useful for later phases. More remote and difficult locations will be easier to bootstrap from small starter sets, because less equipment needs to be imported.

Phase 1 also develops experience in other areas. One is creating a distributed network of individuals and small groups, with equipment in multiple places. As a coordinated effort to produce useful products and services, it can still be considered a factory or business enterprise, just not a traditional one in a single physical place. Another area is building the knowledge and skills of the people using the equipment. Small-scale production isn't taught to most people in their general education, so along with seed equipment and software, we need education and training for the users. Knowing what skills and training is needed will be useful in later phases.

We expect Phase 1 operators to mostly use their skills and machines to make items for themselves, for each other, and for their immediate community. Commercial-scale production with significant sales to the public would fall to the next phase. Since the starter sets are designed for upgrade and expansion, the evolution to the next phase can be natural and gradual. Within this phase, new network nodes and new locations can be started with partial equipment sets sent from existing ones, or purchased from outside.

Phase 2: Distributed and Industrial Development

As more people, their skills, and equipment accumulate in a given location, they can start to sell and trade beyond their own community. They can then get most of their financial support from network operations, where in Phase 1 it was more hobby and part time levels of effort. To the extent people can support themselves, it relieves conventional risks of job insecurity and displacement by automation. In less developed regions it would mean an improved quality of life. These are useful ends in themselves, besides working towards later program goals.

The direct bootstrap approach is to use existing machines to make parts for larger machines. This scales the equipment from home and hobby, to small business, commercial, and industrial sizes. More extensive use of the equipment puts an emphasis on higher duty cycles (percentage of the time it is in operation) and service life (total hours of operation). Where light-duty equipment is adequate at a hobby level, heavy-duty equipment is preferred at industrial levels. Both scale and intensity of use require modified designs from Phase 1, and will require continuing R&D to develop. At smaller scales it is feasible to gather the full range of fproduction machines in one place, and make a wide range of products. At larger scales, the equipment and their operators become more distributed and specialized, and serve larger markets. Location designs and production flows are therefore also modified to account for the changes.

We therefore divide Phase 2 into two sub-phases based on scale and specialization. Not everyone will upgrade to larger scales and intensity of operation, so Phase 1 continues to operate, and there would be a mix of small, medium, and large equipment. Phase 2 locations are in the same moderate environments as Phase 1, with developed and populated areas nearby.

Phase 2A: Distributed Locations - This covers locations that move beyond hobby and home improvement levels of effort within the network, to small business and commercial levels capable of providing most of an individual's financial support, and a wider total range of products and services provided to people outside the network.

Phase 2B: Industrial Locations - These serve larger and more widespread markets at the most efficient levels of size and specialization. This scale may require outside funding for land and equipment, where earlier phases could be self-financed through internal growth. This is because buying and developing an industrial-scale site is hard to do in small increments. Because sources of supply and customers are more widespread, transport capacity becomes more important. Outside market forces also become more important than the internal needs of the community. One way to deal with market forces is distributed ownership across multiple industries, so that people's work and equipment can be reassigned as needed.

Phase 3: Other Earth Development

Difficult and extreme regions on Earth are those which are more than 10 and 20% beyond moderate conditions in any of a number environment and development parameters. See Section 5.2 **Environment Ranges** for details on the parameters. These areas include remote, unpopulated, and undeveloped areas on the surface, such as oceans, deserts, and ice caps. They also include areas away from the surface, such as deep underground below existing cities, or at higher altitudes. These may be physically close to moderate surface locations, but harder to build in.

The more difficult and extreme the conditions, the more that designs have to be modified to accommodate them. This is accomplished by either modifying the working equipment, or artificially moderating the local operating environment to a suitable range. Because of the scope of design modifications needed, we divide this phase into two sub-phases, based on the amounts required. Because conditions can vary in many ways from moderate ones, we expect a number of custom designs will be needed. These locations are generally undeveloped and unpopulated. So starter sets and other equipment can't be obtained locally, and would be delivered from previous locations. They would not typically be small scale, because local operators don't have other jobs to support themselves, or the extra space to house them. The higher costs of transport, and supporting on-site people and remote operations, also tends to make small scale equipment uneconomic.

This major phase allows us to extend civilization to the 6/7ths of the Earth's surface we don't yet significantly use, and vertically both down into the ground and oceans, and up to higher altitudes. The world's population is expected to continue growing in the 21st century. People in currently less-developed areas also want to reach higher levels of economic development. Expansion to new areas will to some degree be needed to satisfy all their needs. This is not so much a need for personal space, since on-going urbanization concentrates people in compact clusters. Rather, it is to access increased energy and material resources. An example project for these areas is to establish greenhouses and water supply systems in deserts to increase food supplies. Automation and remote control allows such locations to operate, even if a lot of people don't want to live there. The two sub-phases are:

Phase 3A: Difficult Earth Locations - These are defined as having at least one parameter 10% or more beyond the moderate range, measured linearly or logarithmically depending on the parameter. The parameters are measured before local development and upgrade. For example, reaching deep underground is difficult before suitable tunnels and shafts are built, but not afterwards. Difficult temperatures are average winter lows below -18C or average summer highs above 42C, such as arctic or hot desert areas. Drier deserts and wet rainforests have difficult levels of water supply - too little or too much. Altitudes above 2750 meters begin to cause problems for some people, and soil strength below 0.19 MPa or ground/water pressure above 2.5 MPa are difficult to build for. Energy supply, primarily from wind and solar, below 125 W/m^2 makes all activities which need power difficult. Locations well below the surface are cut off from these sources. Gravity is nearly constant on Earth, and most places are below 17 mSv/year in background radiation, and are not difficult in these parameters. Communications round-trip (ping time) over 100 ms and normal one-way travel time for people that takes over 2.5 days are considered difficult. Average residence times below 5 years and cargo transport energy above 2.85 MJ/kg impose extra transport burdens from staff turnover and difficulty of outside supply.

Phase 3B: Extreme Earth Locations - Extreme locations use the same parameters, but are at least 20% beyond the moderate range, up to the limits of existing technology. So average daily lows below -23C or highs above 47C are extreme, and only found in severe climates, at high altitude, or deep underground. Water supply below 0.12 or above 3.8 meters/year, and air pressure below 70 kPa or above 120 kPa (+5500m and -1600m altitude) are also extreme conditions. Soil strength below 0.12 MPa, which includes water surface of no strength, are extreme conditions to build on. Therefore all of the ocean surface where it is too deep to build from the sea floor is considered extreme. Ground and water pressure above 3 MPa require extra structural support, and are found at depths below 300 meters in water and 120 meters in rock. Gravity is never extreme on Earth, but some high background radiation areas and high altitudes over the magnetic poles exceed the 21 mSv/year level and are considered extreme. Ping time over 125 ms occurs if modern communications are absent, and travel times over 3 days if conventional transport isn't available. These fall into the extreme range, as are residence times below 3 years 4 months, and cargo transport energy above 3.5 MJ/kg. Determined mining should be able to reach depths of 9.5 km in continental crust and 7 km in oceanic crust in addition to water depth. Counting the oceans, the equivalent of 4.6 billion km^3 of total resources should be accessible.

Phase 4: Orbital Development

We already explore and use space beyond the Earth to some degree. That use is limited by the difficulty and expense of lifting everything needed up the Earth's deep gravity well. We get around this problem by exploiting the energy and materials already in the Phase 4 regions to make things locally. This allows much fuller development of the regions and expansion of civilization. Production capacity from the previous phases is used to build things like rockets and launch sites, which can deliver starter sets and other equipment to orbit. These bootstrap local production, habitation, transport, and services in a progression of locations, starting near Earth, then moving into "open space", away from strong gravity wells. These locations share high levels or full-time sunlight for energy, low gravity for ease of moving large masses, and vacuum which can enable certain production methods. These shared conditions lead to shared designs, so they are grouped into one major phase.

With greater distances from Earth and the Sun there are increasing transportation needs, and thermal, power, and other design changes needed. This leads us to identify six sub-phases by region. Each of these leads to the next in sequence, with products from one phase used to help deliver and set up locations in the next. Orbital production would first support existing space industry, which amounts to nearly 1500 active satellites and $340 billion in economic activity as of 2016. As bootstrapped production lowers costs, current markets should expand, and new one become economic. So, like previous phases, this phase should be self-supporting once started.

Phase 4A: Low Orbit Development - We define Low Earth Orbits as extending from 160 to an average of 2700 km above the surface. The lower bound is set by significant atmospheric drag, which prevents stable orbits. The upper bound is half the potential energy from the lowest orbit to Earth escape. It is stated as an average altitude, because elliptical orbits are possible, which vary constantly in altitude, but not in total energy. Most low orbits are in the Earth's shadow from 22-40% of the time, which lowers available solar power, and typically requires power storage for the time in shadow. Temperatures and lighting are moderated by the nearby Earth, and communications and travel times are relatively short. The natural environment includes the inner part of the **Van Allen radiation belts**, and a modest flux of meteoroids. Material resources include the upper edge of the Earth's atmosphere, and artificial space debris. Other materials would have to be imported from elsewhere.

Phase 4B: High Orbit Development - High Earth Orbits extend from 2700 km average altitude to the limit of the Earth's gravitational dominance at about 1.5 million km. Although this is a large range in distance, it only represents the upper 25% of energy between the Earth's surface and escape. It excludes distances within 35,000 km of the Moon and the Moon itself, which are assigned to Phase 5A below. High orbits are in sunlight 78 to 100% of the time, and temperature is mostly governed by the Sun. Communications and travel times at the outermost edge can extend to 10 seconds ping time and 7 months transit by the most efficient route. Like lower orbits, the natural environment includes high radiation levels from the remainder of the Van Allen belts, and, outside the magnetosphere, from solar and galactic radiation sources. Meteoroid flux is similarly modest, as it is for most of the rest of the Solar system. Material resources in place include a smaller amount of artificial debris, but high orbits are fairly accessible to the Moon and Near Earth Asteroids. Materials can be delivered from these locations, and then the high solar flux used for local production.

Phase 4C: Inner Interplanetary Development - This region includes orbits from as close to the Sun as technically possible to an average of 1.8 AU, where the Main Asteroid Belt starts. It excludes the four major planets Mercury, Venus, Earth, and Mars and the gravitationally bound regions around them. Inner interplanetary orbits are in sunlight 100% of the time, and solar flux varies from 31% to many times that near Earth. Even the lowest levels match the best places on Earth, because there is no night or atmospheric absorption to reduce it. Temperatures vary with amount of solar flux, requiring sunshields or insulation to moderate them for people and equipment. Communications time is up to 1 hour round-trip across this region, and travel time can be several years by lowest energy transfer orbit. There are over **17,000 known objects** in this region as of late 2017, growing about 1500/year. The largest member alone has an estimated mass of 17 trillion tons. So material and energy resources are widely available for production and other purposes.

Phase 4D: Main Belt and Trojan Development - This region includes orbits averaging from 1.8 to 5.2 AU from the Sun, except for the part within 20 million km of Jupiter. Conditions are similar to the previous phase except that solar flux varies from 31 to 3.7% of that near Earth. At the outer reaches, solar reflectors or other power sources such as nuclear may be needed to maintain temperature and supply power. Communications times can range up to 2.88 hours in the worst case, and travel can take 12 years by the most efficient route. Material resources include **over 700,000 known objects** in the Main Belt, Hilda, and Jupiter Trojan groups. Their combined mass is on the order of 3 billion billion tons, all of which is accessible to sufficiently determined mining operations. These materials can be sent slowly but efficiently to inner regions by means of electric propulsion and gravity assists from the major planets. The farther part of the region contains large amounts of water ice and other volatile compounds because of the low temperatures.

Phase 4E: Outer Interplanetary Development - This region includes orbits from 5.2 to 50 AU from the Sun, except for the regions around Saturn, Uranus, and Neptune. Solar flux is quite low in this region, from 3.7 to 0.04% of that near Earth. Nuclear power sources or very large and lightweight solar reflectors will be needed to supply power and keep warm. Without those, objects will naturally be at temperatures of -56 to -200 C, depending on color. Round-trip communications time ranges up to 1.15 days, and travel times would be up to 350 years by the most efficient orbits. This is too long to make economic sense, so actual travel times will depend on the availability of faster transport methods. There are about 1750 known objects in the Kuiper Belt beyond Neptune (30-50 AU), and another 340 Centaurs and short-period comets, whose orbits cross one of the gas giants. This includes Pluto and several other dwarf planets, with a total mass in this region of 4-10% of the Earth's. There are likely many more objects in the region which are too small to find at present. The region is rich in water and frozen gases due to the low temperatures.

Phase 4F: Scattered, Hills, and Oort Development - The final orbital region is the vast one extending from 50 to 100,000 AU average distance from the Sun. This region shares extremely cold temperatures and being close to Solar escape energy. We divide it into three parts by distance - The Scattered Disk from 50 to 2000 AU, Hills Cloud from 2000 to 10,000 AU, and Oort Cloud from 10,000 to 100,000 AU. The outer limit is where the Sun's gravitational dominance ends. Our ability to detect objects in this region is poor at present. This is due to distance, and lack of sunlight to reflect back to us. Only about 335 objects in fixed orbits are known, and a small number of long-period comets. We expect many thousands more await discovery, including a possible planet the size of Neptune. Total mass in this region is not known, but may be a large multiple of the Earth's. Travel, communications, and powering of equipment would be very difficult with current technology. So significant development will await future improvements in those areas.

Phase 5: Planetary System Development

Planetary system locations differ from orbital ones in being tied to relatively large gravity wells, requiring additional transport capacity to traverse. They can also experience shadowing or night when close to or on the surface, and higher radiation levels from trapped particle belts. The surfaces of large bodies have significant gravity levels, and sometimes an atmosphere. All of these conditions are different from those in the open space of Phase 4, leading to a different phase with different designs for them. The various planetary systems also differ from each other, requiring designs to accommodate each. So we provide five sub-phases to cover the range of such locations. The sub-phases are in approximate order of difficulty. The start of sub-phases for Orbital and Planetary development overlap in time. Each planetary one starts after the orbital regions which must be traversed to reach them.

Phase 5A: Lunar Development - The Lunar region includes the Moon itself, and orbits within 35,000 km on average from the Moon's center. The Moon is relatively close in physical distance to Earth, but reaching the surface requires significant additional energy due to it's gravity well. Therefore it comes after Phase 4B: High Orbit in terms of difficulty. Communications and travel times are relatively short, less than for some parts of High Orbit. Available sunlight per area of the Lunar surface is 50% of that for high orbit on the equator, and less at higher latitudes. Availability can reach 99% for the highest circum-lunar orbits. Average temperatures are similar to that on Earth, but exposed areas can vary hundreds of degrees between light and dark. Determined mining on the Moon should be able to reach depths of 50 km before rock pressure makes it unreasonably difficult. This makes 2 billion cubic km of resources potentially available. Because of its origin and history, the Moon is depleted in **Volatiles** compared to even nearby asteroids. These are elements and compounds with relatively low boiling points, such as water and atmosphere found on Earth. Conversely, **Lunar Geology** indicates a high percentage of metallic oxides, making it a good source of oxygen and various metals.

Phase 5B: Mars Development - The Mars region includes the planet itself, and orbits up to 340,000 km in average distance from its center. This includes the natural moons Phobos and Deimos. The orbit of Mars is eccentric, so solar flux varies from 36 to 52.5% of the near-Earth reference amount. Surface gravity on Mars varies from 3.68 to 3.74 m/s^2 (37.5% of Earth), and atmospheric pressure varies from 30 to 1155 Pascals (0.03-1.14% of Earth). Mars locations would follow phase 4C: Inner Interplanetary, since you must travel through interplanetary space to reach Mars. Determined mining should be able reach depths of about 25 km, providing 3.6 billion cubic km of accessible resources, nearly double than from the Moon. Mars retains significant water, and has a quite varied geology.

Phase 5C: Venus and Mercury Development - These regions include the two planets, and orbits averaging less than 600,000 and 100,000 km from their centers. A relative lack of known nearby asteroids or moons, higher delta-V to reach their orbits and land on them, and extremely hot or high pressure conditions places these locations after Mars. Their advantage consists of 1.9 and 4.6-10.6 times more solar flux, providing ample energy to extract and process resources. The lack of known asteroids in this region may be due to the short orbit times, which allow frequent planet flybys. A given asteroid will then be removed by either hitting a planet, shifting the orbit outwards, or shifting it so close to the Sun it vaporizes. New objects are supplied by gravitational changes from farther orbits.

The lack may be partly due to the difficulty of finding small objects when looking towards the Sun. If the planets themselves and nearby asteroids prove insufficient or the wrong composition, it can be made up by imports from the better supplied outer regions.

Phase 5D: Jupiter System Development - The Jupiter system includes the largest planet, 69 known moons, four of which are very large, and orbits within 20 million km of the planet's center. The larger moons are useful for gravity assists, making travel between locations easier, but the great mass of Jupiter makes reaching the planet itself very hard. Development of Jupiter would logically follow Phase 4D: Main Belt and Trojan. Jupiter sits between it's Trojan clusters, and the outermost moons are really loosely captured asteroids, making it an easy next step. Progressing inwards needs more transport energy, and you encounter lethal radiation levels, requiring lots of shielding for people and equipment. The large moons represent many billions of cubic km of resources, which would eventually make them attractive sources. However solar flux of 3.3-4.1% of Earth's is a challenge to supply enough power to make use of those resources.

Phase 5E: Outer Gas Giant Development - This phase includes Saturn, Uranus, and Neptune, and the surrounding regions within 20, 12, and 12 million km respectively. It includes 103 known moons, a number of which are large, and three ring systems, one of which is famously prominent. Solar flux is 1%, 1/4%, and 1/9% of Earth's, respectively, making alternate energy sources very attractive. These locations follow Phase 4E: Outer Interplanetary, as you must travel through that region to reach the three planets and their surroundings. This phase is far enough in distance and needed technology that a significant amount of research and development is needed before it can be used.

Phase 6: Interstellar Development

The last major phase includes locations which are not tied to the Sun's gravity well. This includes open interstellar space not tied to a specific star, and the regions around other stellar systems. There is no reason to stop the expansion of civilization at the borders of our Solar System, assuming we have the necessary technology and it is economically reasonable. However, at present we don't have full knowledge of what planetary systems exist around even the nearest stars, and transport technology to reach them in reasonable time is currently speculative. We include this phase as a long-range program goal, but with the understanding that most of the details will have to wait. We divide this phase into three sub-phases, to account for the different environments and activities in nearby interstellar space and exostellar systems, and for farther reaches of the Milky Way.

Phase 6A: Nearby Interstellar Development - This region begins 100,000 AU from the Sun, where it's gravity is no longer dominant. For design purposes we set an arbitrary limit of 20 light-years from the Sun. If we can reach that distance, and restock/rebuild our equipment, then later projects can travel further in increments of 20 light years, but not require new designs. The interstellar region excludes the stars within 20 light-years and their respective regions of gravitational dominance. They are assigned to Phase 6B since they have more in common with previous phases closer to the Sun than the spaces between stars. The volume of this region therefore resembles the solid portion of Swiss cheese, with scattered holes that are not included. Contents of this region are poorly known as of 2017, but very low density. It includes the interstellar medium of gas, dust, particles, and radiation. It likely includes a population of objects larger than dust but smaller than stars, which are so far mainly undetected.

Phase 6B: Nearby Exostellar Development - There are currently about 105 known **Stellar and Brown Dwarf** star systems within 20 light years. We know much more about the stars than what orbits them or the spaces in between, because stars are bright and we can more easily collect information from their light. Each star has a region of gravitational dominance over the Galaxy as a whole and other nearby objects. This is estimated to be 100,000 AU times the square root of the system mass in units of our Sun's mass. The stars are all in motion relative to each other, so the population within 20 light-years changes on average every 1150 years. There are about two dozen known planets in this region, and two have circumstellar disks, but this data is incomplete due to the difficulty in detecting planets. Design for these locations must await better information, and a lot of new technology development.

Phase 6C: Farther Interstellar Development - Our last sub-phase is a place-holder to cover the remainder of the accessible Universe. It begins 20 light-years from the Sun and extends as far as transport methods make it possible to reach. Since current and near-term methods are far from able to reach such distances, work on this sub-phase beyond transportation improvements is reserved to some point in the future.

Program Structure

To make a large and complex program comprehensible, we divide it into multiple levels of detail, with each level dividing a given part into a reasonable number of smaller parts. At the most detailed level, individual program elements are small and simple enough to be designed and implemented, without further division. For this example of a complex program we will not carry it to that level of detail. It would first require completing conceptual and preliminary design, which isn't done yet, and there isn't space in this wikibook to include all the information. Instead, we will present the top several levels and the process of defining their elements, with the understanding that future work can use similar processes at the lower levels.

Level 1: Program

The top level is our program as a whole. It is here that overall goals and objectives are defined, and top level interactions occur with the rest of civilization. Decisions are also made to implement parts or all of the program vs. staying with what is already in progress and planned. Change takes more work than staying with what already exists. So a new program has to be sufficiently better to motivate people to change. That requires developing the needed ideas and technology, and educating people about why they are better. We refer to what already exists as the **Existing Baseline**. For example, in space launch that would include rockets that are already in operation, and new ones already funded and in development. For factory production it would include currently operating factories, and the state of the art for building new ones. Our program is not yet developed enough to recommend implementing it, but we think it has great potential. This is what has motivated our work to date, and we hope to continue it until we can make a recommendation for or against it.

Level 2: Phases

Level 2 of the program includes the major program phases and sub-phases described above, and in the later sections of Part 4. These phases inherit parts of the top level goals and objectives, such that all the phases together meet all of them. Earlier phases and their parts may only meet some of the goals, or at lower levels of performance, with the intent to upgrade and expand the coverage later on. Part of the early design process is to specify what goals and performance levels each phase will accomplish, and timing for when they reach them. Since the phases are covered in more detail elsewhere, we won't duplicate that information here.

Level 3: Projects

Each phase can include one or more projects. These projects are intended to accomplish a task or meet a goal. For example, the **Seed Factory Project**, which has already been started, is intended to design and test prototype systems to prove out ideas for self-expanding production. This project is part of Phase 0: R&D. A hypothetical "Floating Cities Project" would be part of Phase 3, due to the difficult or extreme environment of the oceans.

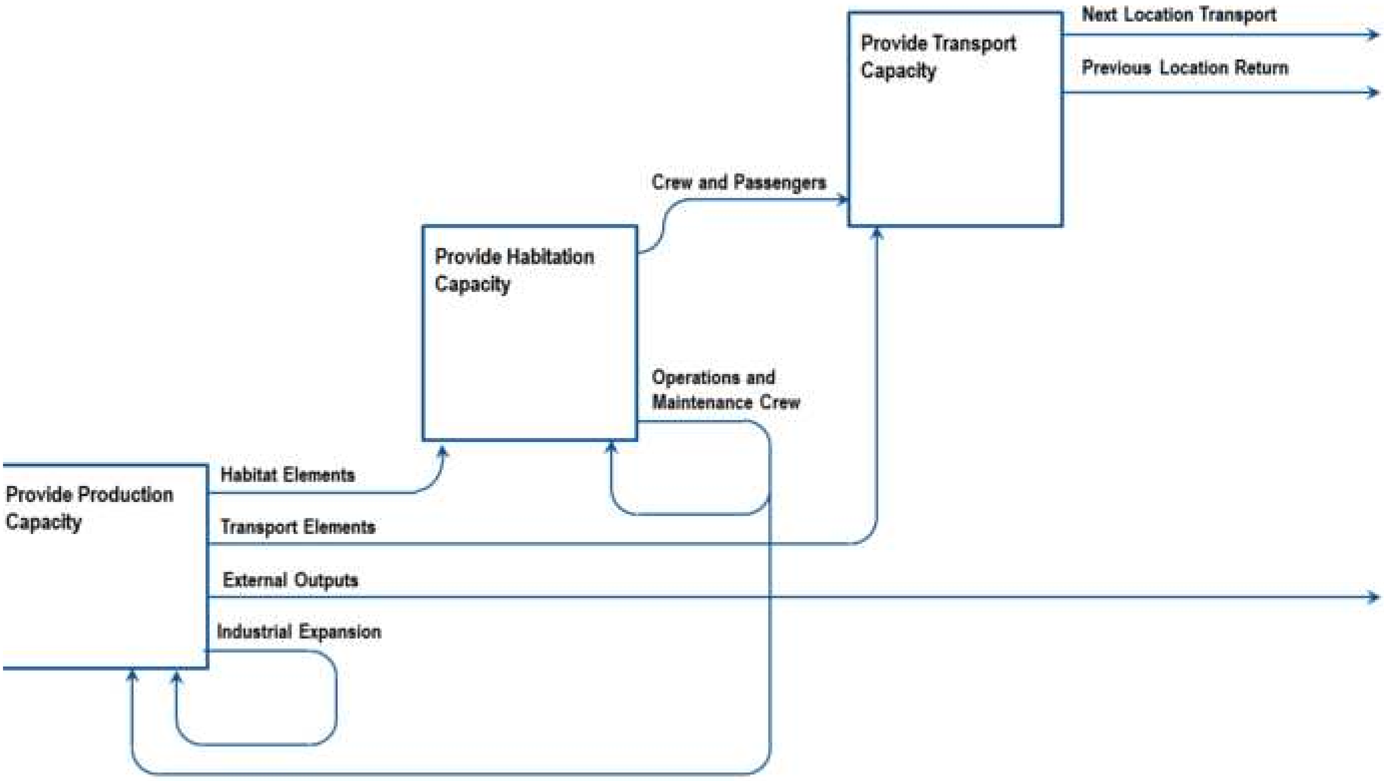
Level 4: Locations

The program is aimed at upgrading and expanding civilization for the reasons listed in the Program Goals section above. The general approach is to build up a given location, adding functional elements that allow internal growth and useful outputs. Each location helps support itself, interacts with previous locations, and with the rest of civilization. When enough capacity has built up at current locations, they can deliver starter elements to a new location. The starter set then repeats the cycle of growth and useful outputs. Parts and materials are delivered from previous locations for whatever cannot yet be made locally. Locations are defined by ease of internal local transport for people and bulk goods. An example is a metropolitan area around a city, where travel times are a couple of hours or less between the various parts, and a good road network exists. Transport between locations will happen, but the increased time and cost will make it less frequent, and preferably for high value, low mass items. So you may ship a computer to a distant location, but you prefer not to send a truckload of gravel that far.

Locations in undeveloped, difficult, or extreme environments, and in space, may be much smaller than a city at first. This is due to lack of transport systems like roads and airports, and large distances to the next location. In developed regions, the overall size of a program location is typically that of a metropolitan area, but does not include all the contents of that area. It only involves program-specific people and equipment located there. They interact with the surrounding non-program region, and other program locations.

Level 5: Functions

The next level below a single location are the functions performed there. All of civilization shares common functions, such as protection from bad weather, supplying food and drink, fabricating parts, and assembling them into useful items. Eating utensils and CNC laser cutters, for example, can be more or less the same from one location to the next. So we don't have to design unique elements for every location. Instead, we can identify the functions that need to be performed there, then copy existing



xample functional diagram.

designs or modify them as needed to satisfy the needs. Each function, like cutting parts from stock material, has inputs like electricity, and instructions on what part to make, and outputs like the finished parts and unused scrap. These inputs and outputs connect functions to each other, and to outside the program boundary. The functions and their connections can be displayed in various ways, such as the example diagram in Figure 4.1-3.

Because civilization shares common functions and these functions share common connections between them, we can define a **Reference Architecture** that applies to existing elements of current civilization, and new elements of our program. This makes it easier to compare what already exists to what is new, and saves us from having to define theorganization of new locations each time.A given location may only include a subset of the common functions, especially when it is new. Locations can evolve over time, adding new functions. The program as a whole also evolves over time, adding new functions and upgrading existing ones.

Further Levels

Functions can be subdivided into more detailed ones, and then systems and specific elements designed to perform those functions. This work falls into the various fields of engineering. For space systems, some of the methods and concepts are described in earlier parts of this book. For other kinds systems, we refer the reader to the enormous range of technical literature for them. For reasons of space and time, we will not cover this level of detail in our current work, but we will note it is eventually needed to implement actual projects.

Section 4.2: Phase 0 - Research & Development

The main program goals are upgrading civilization on Earth, and progressively expanding to more difficult environments, including space. To accomplish these goals, some new or improved technologies and methods will be needed. Once available, they can be incorporated into suitable designs for their intended locations. For space locations in particular, there has been a severe lack of production and habitation capacity, with the main focus so far being on transport and information services. This imbalance exists to a lesser degree in difficult Earth environments. For example, many ships (transport) cross the oceans, but relatively few things are produced there, and few people inhabit the seas. Phase 0 is therefore included in the program as preparation for what's needed for the later phases, and to consciously correct unbalanced development where possible. The major goals of Research and Development (R&D) phase can then be stated as:

- (1) Identify systems and elements covering the full range of production, habitation, transport, and services functions
- (2) Supply needed new technologies and methods, in the form of tested and ready to use elements.
- (3) Supply detailed designs for equipment and locations, incorporating both existing and new elements.

R&D Planning

The program as a whole is complex. **Systems Engineering** methods (see Section 1.5) have developed to handle such complexity, so we intend to use them to for this and later phases. Other engineering methods will also be used where appropriate, but the systems approach is especially useful across whole programs. This includes their interactions with the world outside the program, and the constituent parts of a program with each other. Part of the systems process is to break down complexity into smaller parts, which are then more comprehensible and easier to design. We have already started this in Section 4.1 by identifying a sequence of phases and sub-phases according to scale, type of environment, and distance. A given set of locations within a sub-phase can then share similar designs, and to a lesser degree with those of the major phase and program as a whole.

Civilization as a whole has common elements across all of it. For example, people need protection from the environment and food to eat no matter where they are. Heat-treating alloy steel can use the same process anywhere you need to do it. We can therefore define a reference architecture for these common elements of civilization, and apply it to organize the tasks of upgrade and expansion. Many existing parts of civilization and component technologies are good enough as-is for what we would like to do. In those cases, we don't have to change them, just use them. Other items are deficient or undeveloped, and not being pursued elsewhere. As we identify them, we can rank them by parameters like benefit ratio, cost, difficulty, probability of success, and time to complete. Then we can add them to our R&D plans in the best sequence. New technology we develop will be used internally in the program, and also supplied as a benefit to civilization at large to use elsewhere. These external uses will also be considered in deciding what R&D to pursue.

The R&D work can be divided into a general part that applies across the whole program, and sub-phases covering work for specific environments and locations. The sub-phases and tasks are detailed further below. Limits on our current knowledge, and on available project resources, mean we cannot do all the R&D work in advance or all at once. In some cases, a given area of R&D must be completed successfully before following work can be done. Since we do not know in advance if we will succeed, we expect that R&D plans will often need to change, or follow multiple paths. We also expect progress across civilization in other technologies, so a given design may no longer be optimal and require upgrades. Therefore the R&D phase is expected to continue in parallel with later phases for as long as the program continues.

Products and services produced as outputs from early phases can be used internally to support later R&D. For example, we may demonstrate self-expansion of an industrial building as an R&D task, and later use that building for further R&D, or as a production area whose product sales finance further work. Field experience from earlier phases can be fed back to the R&D phase to improve later designs. Self-use and feedback should also be considered in R&D planning.

R&D Process

The R&D process will be similar across all phases and subphases. We give the component tasks a consistent numbering system so they may be coordinated and compared across the program:

Task 0: Coordinate R&D - This activity includes coordinating external resource flows, tasking, planning and scheduling, and analyses.

0.1: Coordinate R&D External Flows - This task includes arranging and managing resource flows into and out of the phase, from outside the program and to other phases.
0.2: Coordinate R&D Tasking - This activity includes arranging which resources will be applied to the phase tasks.
0.3: Coordinate R&D Planning & Scheduling - This activity includes developing future plans and schedules for the phase.
0.4: Coordinate R&D Analyses - This activity examines past performance and studies ways to improve R&D processes.

Task 1: Conceptual Design - This includes exploring new concepts and developing a reference architecture. This is followed by a systems engineering process to reach a concept level design. This includes defining the main functions and elements of the program, and how they will be operated and maintained across the stages of their life cycle. This model is itself part of the conceptual design. Based on prior experience, systems engineering effort optimizes the program cost and schedule at ~10-20% of total effort, with the systems tasks weighted towards the early part of the program. The systems engineering process flow is used iteratively in later design stages. The subtasks here are a template for those flows, but to avoid repetitiveness they are not broken out separately each time below.

1.1: Explore New Concepts - This step covers taking ideas, such as self-expansion and automation, and applying them to create new products and projects. Some concepts only apply to particular phases or elements, so an application matrix is an output of this task.
1.2: Develop Reference Architecture - The reference architecture is a high level design used to identify technology risks and readiness level (TRL), and make early estimates of cost and schedule. It is a starting point for the conceptual level design. It includes program goals, an architecture description, high level interfaces, element requirements, and element descriptions. Supporting data for the reference architecture includes data sources, analyses to support concept selection, and tracking from goals to lower elements.
1.3: Identify Requirements & Measures - These establish measurable features a design must meet, and criteria for selecting among design alternatives. See Section 1.5 **Requirements Analysis** for details.
1.4: Perform Functional Analyses - This breaks down what the design does in terms of functions it performs or a sequence of operations. See Section 1.5 **Functional Analysis** for details.
1.5: Allocate Requirements - This assigns the requirements from task 1.3 to functions from task 1.4 to ensure they are all met somewhere in the design.
1.6: Model Alternatives & Systems - There are many possible ways to meet a given set of requirements. Modeling the options provides measurable details for each. The modeling process includes

(1) Collect External Technical Information: This includes data needed for modeling and later design, such as existing product data, industry specialist contacts, and current state of the art such as books and articles.
(2) Develop Alternative Options
(3) Build System Models.

1.7: Optimize & Trade-Off Alternatives - This includes varying parameters of a design option, and comparing different options, to find the ones that best meet the selection criteria.
1.8: Synthesize & Document Design - The outputs from this task are articles, reports, and books documenting the chosen concepts.

Task 2: Preliminary Design - Assuming the conceptual design produces a sufficiently promising concept, the next stage is to define the elements of the program in more detail. This is done in parallel with component technology (task 4) and prototype systems (task 5) because otherwise size and performance would be too uncertain. Multiple design alternatives may exist in this stage, until competing technologies and testing is far enough along to permit selection. This follows same steps as conceptual design, but at greater level of detail.

Task 3: Build R&D Locations - This activity includes building or acquiring use of offices, research workshops, conventional production shops, and prototype test areas needed for R&D work. The design requirements for the R&D locations comes from the previous design work, needs for technology development, and for building and testing prototypes. Consideration is also given to adapting the R&D facilities for later phase use.

Task 4: Develop New Technology - This includes identifying the performance needed based on the conceptual and early preliminary design, surveying the status of current technology, ranking areas for improvement in terms of impact, then applying effort in the most promising areas to improve performance or lower uncertainty. Some technologies are already under heavy development outside the program. So rather than duplicate that effort, we select areas where a limited budget can have the most impact, or encourage others to invest in those areas which need the most work. Technology level work is aimed at single processes or components.

Task 5: Build Prototype Elements - At some point it becomes necessary validate integrated elements and demonstrate performance levels by building prototype hardware. This can be simplified versions of what will become final designs, scaled versions for what will be larger designs, or versions that demonstrate the functionality, but do not use the final materials and components because they have not yet been made. Prototype elements may carry over to later phases if they work well enough, or can be upgraded to final versions in some cases.

Task 6: Test Prototypes - This task reduces technical risk by demonstrating the actual performance of prototype system elements. Initial testing would use the local R&D environment conditions, but later testing uses the full range of operating environments, either in test chambers or by taking the equipment to suitable locations. Deficiencies found during testing are fed back to developing new technology. Early prototypes of a given element may have lower performance goals, which are later increased as improved designs are developed. In some cases, test units can go on to be used operationally, and therefore transferred to a later phase.

Task 7: Detailed Element Design - This activity covers detailed design of specific facility locations and equipment. This includes places places where R&D is done, operational facilities for later phases, individual equipment items within these locations, and vehicles and other equipment which moves within and between locations.Detailed designs incorporate existing technology, plus new technology developed, prototyped, and tested within the program. They can also include off-the-shelf equipment, parts, and materials from outside the program when that makes sense. Because of improved technology over time, goals for further expansion and upgrade of existing systems, and development of new locations, this task is expected to continue through the program. Particular designs for complete facilities, processes, equipment, vehicles, and components can be used in multiple projects and phases, or sold as separate products.

R&D Sub-Phases and Tasks

The sub-phases, and the R&D tasks identified for them so far, are listed below. This list is preliminary, since concept exploration for the later program phases is incomplete. The tasks are listed in the order we identified them, rather than time order, since determining the best sequence for the work and schedule planning is a later step. For identification we use the plain "Phase 0" label, with no additional letters, to identify general R&D work which applies across the whole program. When the work is specific to a single phase, a letter is added, such as 0A or 0B. When the R&D work applies to two or more later program phases it is identified with all applicable sub-phase letters, thus 0CD or 0G-L. Since some R&D locations may themselves need new technology and design, the first lettered sub-phase, 0A, applies to Phase 0 itself.

Phase 0 - Program-Wide R&D

The program's goals are to establish new locations for civilization, and upgrade existing ones. We have identified self-expanding systems, using **Seed Factories** as starting points, as a key technology to reach those goals. It can be applied to existing locations across current civilization, and to new ones both on Earth and in space. The general approach of self-expansion includes more specific methods like distributed production networks, remote-controlled operation, and smart tools which can operate themselves. Manufacturing in general, and automation in particular, already get a lot of engineering effort, so we do not need to duplicate those efforts. Our R&D work will focus on the unique aspects of self-expanding systems, and integrating other technologies into them.

Self-expanding systems, seed factories, and the related ideas fit within the more general subject of **Advanced Manufacturing**. They can be used anywhere, including both developed and undeveloped regions on Earth. However, the main subject of this book is space systems. So we devote a separate book entitled **Seed Factories** to general discussion of those ideas, and their applications on Earth. We provide a short introduction here, and mention such systems where they are used in later phases.

Seed Factories Introduction - All factories produce products, and some factories produce the same kinds of products they use themselves. For example, a **Steel Mill** typically uses some steel in its own construction. Self-expanding factories are specifically designed to use their own output to grow. A "seed factory" is an optimized starter set of people and equipment. It includes plans and instructions for a chain of expansions to reach some desired mature state. It may also include a starting inventory of materials and parts to use in production. Using tools to make more tools is not a new idea. In fact it is nearly as old as tool-making itself. What is new is optimizing for a small starter set to bootstrap the process, applying modern computer systems, automation, robotics, and AI to the task, and combining several growth paths to increase output:

- Making identical copies of the starter equipment,
- Making larger versions of, or extensions for, the starter elements, and
- Making new tools and machines not in the original starter set. These can be used for new tasks, and expand the range of possible outputs.

The present state of technology does not allow for full artificial **Self-Replication**, which means copying 100% of its own parts and doing so without help from people. Self-expanding systems therefore include more than just the factories that make physical products. Whatever parts and materials a factory cannot make itself have to be supplied from elsewhere, therefore transportation is needed. The people required to run the factories have other needs and desires beyond their work. They also need places to live, food, and a variety of services. In developed areas much of that is already available. But in undeveloped parts of the Earth, and in space, complete self-expanding systems will need production, transport, habitation, and services in order to grow. Unlike current machines, people and other biological systems can copy themselves. So they can grow as needed within a complete self-expanding system.

The seed factory approach should be worth developing on Earth for its own sake. It should make setting up new factories cheaper, especially in remote or difficult locations, because you only need a starter set and not the full factory. With an emphasis on self-growth, they may also achieve high economic rates of return. Once developed on Earth, industrial-scale factories can produce items needed to reach space, such as launch sites and rocket factories. New starter sets can then be delivered to space locations, and the expansion process continued. The experience gained on Earth, and the leverage from a series of self-expanding factories, multiplies the savings on future space projects, making them much more affordable.

Phase 0A - R&D for Phase 0 R&D Locations

Even when no new equipment designs or technology is needed, research and development for later phases will need offices, laboratories, prototype fabrication, and test sites. We therefore apply the R&D process above to design and build these locations. When new and unique items are needed, such as a special test chamber, they are developed and built the same way as other equipment for the later phases. Specific needs for this phase have not yet been identified.

Phase 0B: R&D for Phase 1 Starter Projects & Network

Starter Sets & Bootstrapping Paths - Work is needed to identify the best starter sets and growth paths for small-scale production, and what new technology and designs, if any, are needed. Some people already have sets of tools they can use immediately. Many other off-the-shelf tools and machines are available at reasonable cost. These may be sufficient to start the bootstrapping process, but some custom designs may be helpful. Plans and instructions that people can follow need to be recorded and distributed, along with whatever training materials are needed for people who don't have the necessary skills and experience.

Distributed Production Networks - Traditional factories and large office buildings brought equipment and people together in one place because it was the only way to efficiently organize the work. Modern communications and transport networks relieve the need to be in one physical place, and allow coordination of distributed work in many places. Some prominent examples are development of open-source software, and **Wikipedia**. In a modern production system, the control of the machinery can be a mix of a on-site people, remote control by people, and automatic control by computers and software. Since all the people don't have to be nearby, you can operate in undeveloped, hostile, or expensive locations more easily, and with less of an environmental footprint. Remote operators can efficiently split and re-assign their work as needed between locations.

It is likely that some machines and workers will still be grouped together in shared locations, for efficiency or other reasons. Modern technology merely removes the requirement that they all have to be om one place. Distributed production is helpful in the earliest stage of small-scale operation, because you can avoid the cost of a dedicated larger site. Instead, people can use spare space where they already live, or temporary work sites when and where needed. Multiple small efforts like these can then combine to finish larger projects. We therefore place initial R&D for distributed production in Phase 1, but can use it in all the later phases.

Some of the needed technology for distributed production already exists. The R&D tasks for this sub-phase are then to improve or fill in the parts that do not, and combine them into flexible distributed networks. The flexibility is needed because the program intends to constantly add new locations, and existing ones will self-expand. So we cannot operate on the basis of static networks. They have to grow and adapt along with the rest of the program.

Applications to Later Phases - In this phase, a goal for distributed production technology is the capability to connect and operate hobby and home improvement level equipment in fairly close proximity, like a single metropolitan area. Later phases would need upgrades for long-distance remote operations, such as on the Moon from Earth, or Mars from Phobos. More R&D may be needed later on in this technical area. Space is a particularly undeveloped, hostile, and expensive location. So when you optimize your operations you will want to minimize the on-site humans, and maximize remote control, and using smart tools which can operate themselves. So at first there will be a strong incentive for the upgraded technology. As factories, habitats, and transportation systems are built for the later phases, people can be supported more easily on-site. So the optimum balance of local people vs remote and smart tools will shift. Having gained experience with the distributed approach on Earth, using it in space will not be something entirely new, but rather an extension of what was learned in earlier phases.

Phase 0C - R&D for Phase 2A Distributed Locations

R&D for this sub-phase involves design of more specialized and larger machines than for Phase 1. These are used for small business and commercial activities, therefore would have higher duty cycles and longer operating lives. Besides design for these conditions, another R&D topic is the best growth paths from the previous phase, and expansion across a wider range of industry categories. A third R&D area is the grouping of varied size equipment in terms of more specialized and distributed sites across a location, and linkages between locations. All of these R&D areas continue in the next sub-phase to the industrial scale, which uses the largest size equipment.

Phase 0D - R&D for Phase 2B Industrial Locations

This sub-phase completes the sequence of growth to larger and more specialized equipment, for developed locations in moderate environments on Earth. It includes equipment for the full range of production, habitation, transport, and service industries. Equipment for all these industries already exists, and is widely produced. The R&D for this sub-phase includes modifying their design so they can be made by self-expanding and distributed systems. It also includes the growth paths and methods to reach industrial scale from the smaller scales in earlier phases.

More specific R&D tasks may be identified later, for particular industry groups or individual industries. One that we know of at this point is industrial transport to Low Orbit, since it will be needed for the later program phases in space:

3. Industrial Transport

Launch to Low Orbits - This is placed in Phase 2B because traditional rocket factories and launch sites are industrial-scale facilities on Earth. Locations for Phases 4-6 are in space, but still interact with civilization on Earth. So there will be a continuing need for transport from Earth to orbit, and back. Obviously space programs already exist, and many satellites are in orbit, but their cost is high. Partly this is due to the transport cost itself, and partly due to lack of production in space. This forces all equipment and supplies to come from Earth. In-space production is addressed in the later phases, while this topic covers transport needs.

In the earlier parts of phases 4-6, transport needs to orbit will be relatively small. They can use existing launch systems, or ones currently in development, to avoid the cost of unique development. As program traffic increases, the advantage of new and more efficient systems will grow. The R&D in this sub-phase will then cover such new systems, beyond those already in development elsewhere. Sections 4.4, 4.6, and 4.7 present some early concepts for this R&D work. In section 4.4 - **Phase 2B Industrial Locations** we consider a small, 3 stage, fully re-used conventional rocket and some other alternatives for the "build our own" option. The design is not complete enough to decide between make or buy yet. The intent is when traffic is sufficient, the start-up transportation will be augmented or replaced with larger, more efficient, and specialized launchers. The initial cargo may consist of assembly robots and parts for an initial orbital platform. If we are building our own launcher we want to make it as small as practical to keep the design and construction cost low

Upgraded Transport to Low Orbit - The program will add upgraded transport when there is sufficient traffic to justify the capital cost. Again, there is always the option to use transport from outside the program, but we consider various internal alternatives using our self-production capacity. On Earth we use different transport systems for bulk cargo than for passengers for cost and safety reasons. One alternative is to specialize our space transport elements for the same reasons. Section **4.6 - Hypervelocity Launcher** presents a high acceleration gun for launching bulk cargo such as propellants or structural parts. Delicate cargo and humans would travel by other methods. The launcher gives the cargo a large starting velocity, so it substitutes for part of the rocket stages. In theory it should lower cost because a fixed gun can be designed to fire many times, and is made from industrial pipeline quality parts, which are much cheaper than aerospace grade parts.

Section **4.7 - Low G Transport** looks at methods for transporting humans and cargo which cannot withstand the high acceleration of the hypervelocity launcher. The choice of which to use depends on results of more detailed design and what other launchers area available outside the program. Some candidates to build our own systems are a combined air-breathing/rocket system, or a gas accelerator similar to the hypervelocity launcher, but lower g level, followed by air breathing or rocket stages. Separate stages will be easier to develop, modify and upgrade than a single integrated vehicle, although there will be a penalty in operations cost. A single integrated vehicle can be developed later once traffic will support the more complex design.

Phase 0E - R&D for Phase 3A Difficult Earth Locations

Difficult and Extreme locations involve all the sizes from small to large that were developed for Phases 1 and 2, but in a different environment. Therefore the existing designs will sometimes need modifications, and in other cases unique designs will be needed. The effort to set up in remote or hostile conditions will tend to make small scale equipment less likely, and the emphasis shift to larger sizes. Example difficult environments include very cold and hot regions, deserts and rain forests, altitudes above 2750 m, weak soils, water and ground depths of 250 and 100 meters respectively, areas of low energy resource or high natural radiation, high communication and travel time, low stay times, and high transport energy, or combinations of these conditions. Each may require R&D to accommodate the particular circumstances.

Phase 0F - R&D for Phase 3B Extreme Earth Locations

Extreme locations are an extension of difficult ones, but farther from moderate conditions up to the limits of technology. R&D would be needed to push technical limits beyond the state of the art. An example would be hard rock mining more than 5 km below the surface, well below the deepest current mines. Some example extreme environments include very cold conditions in parts of Antarctica, The open ocean surface, which has zero ground strength, great depths underwater or underground, and the most remote and inaccessible surface locations.

Phase 0G - R&D for Phase 4A Low Orbit Development

Low Earth Orbit already hosts many satellites, and as of the start of 2017, two space stations with a total crew of eight. However it lacks significant production capacity, aside from assembly of pre-made elements at the stations. Eight people is only one billionth of the Earth's population, and no transport systems are based in low orbit. What transport exists is all based on Earth. So while we have a foothold in low orbit, civilization can't be said to have fully expanded to this region. The R&D for this sub-phase is then aimed at full use of low orbits, beyond current programs and activities. So far we have identified the following:

1. Low Orbit Production

1.2. Supply Power - Electrical power using solar panels and batteries is fairly well developed for low orbit. Sunlight is available at least 60% of the time, but only special low orbits have it all the time. So energy storage is needed to bridge the time in the Earth's shadow. Thermal power using solar concentrators is an area for R&D.

1.3. Extract Materials - Low orbit has two significant sources of materials besides those brought from Earth or more distant locations. The first is the upper fringes of the atmosphere, which can be collected by an orbital compression scoop. The second is space debris, which includes non-functional satellites, empty stages, and collision fragments. Some of the gathered gases can be used as propellant for collecting the debris, since they are in widely scattered orbits. The space debris at the least is a hazard, and removal is a benefit to other space activities. But it consists of aerospace-grade parts and materials, some of which may still be functional. Salvage and recycling of these items would save having to launch comparable items from Earth. R&D is required to prove the gas mining, collection, and reuse of old hardware is practical. It would also provide some experience for later mining and production beyond low orbit.

1.4 Materials Processing & 1.5 Parts Fabrication - Very little of either of these has been done in orbit and in zero gravity. Extensive experiments and prototyping are needed to find out which terrestrial methods can be used, how they may need to be modified, and what new methods can be used in the unique orbital conditions.

1.7 Low Orbit Assembly - The design of transport systems typically is much more expensive than a single use of them. Therefore a number of deliveries on a smaller launch system is preferred on cost to a single delivery on a very large one. This in turn drives a need for assembly of larger elements in orbit. **Section 4.5 - Orbital Assembly** gives one approach, using an assembly platform in low orbit. At first, the platform assembles pre-made components launched from Earth. As other production elements get added, it later shifts to assembling a mix of Earth and locally made items. The first task of the assembly platform would be to bootstrap its own construction. The platform is then used to assemble larger payloads, and then then later build seed elements and vehicles for new locations. Humans are kept to a minimum in the early stages because of cost. The assembly robots start out mostly controlled from the ground. Some experience already exists with orbital assembly of space stations, and similar maintenance and repair tasks for the Hubble Space Telescope. The R&D tasks here are whatever improvements are needed beyond these levels.

2. Low Orbit Habitation

Partial Gravity Research - There have been a series of space stations in low orbit, which provide experience in zero gravity conditions. For people, at least, long periods of zero gravity (up to a year) are detrimental to health. We have essentially no information on what gravity levels between 0 and 1g do to people and other living things. Therefore we don't know what designs are needed for long-term habitation or agriculture. Besides living things, some production methods work better with gravity, but the minimum required isn't well understood. A Variable Gravity Research Facility would start to answer these questions by providing adjustable artificial gravity. The design can include one or more modules on a rotating arm, and their position and rotation rate adjusted to get desired gravity levels. An alternate course is to assume full Earth gravity as a design requirement at first, then pursue partial gravity research on an "as available" basis. For example, a rotating habitat which produces full gravity at the rim will have areas of lower gravity that can be used for research. Research stations on the Moon and Mars can provide data on their particular gravity levels, limiting stay times for people to less than a year at first, until more experience is gained.

Habitat Growth and Upgrade - Habitats will generally start small and grow over time. So another research area is the best growth paths for the them: in physical size, from possibly zero-g to some gravity level, from open food and air cycles to closed life support, and from hardware supplied from Earth to local production. The design of the habitats is likely to be complex, and we can only lay out these open questions as a starting point for further R&D work.

3. Low Orbit Transport

This category covers transport that operates within low orbits, and reaches farther destinations. Transport to low orbit is covered under Industrial or Difficult Earth Locations, because that is where they are built and start from.

Electric Propulsion - Ion and plasma engines have about 5–10 times the fuel efficiency of conventional rockets, and have already seen some operational use. Section **4.8 - Electric Propulsion** looks at options for propulsion modules, which can be used singly for smaller missions and in multiple units for larger missions. There are several types of electric engines available, but they will be needed in some form if missions beyond Earth orbit are to be done economically. The higher efficiency allows bringing the vehicle back and using it multiple times, a key cost savings. R&D for this phase would be aimed at upgrading the propulsion to higher power levels, and enabling use of mined propellants rather than the scarce Xenon used today.

Electric propulsion can be used within low orbits for drag makeup, for changing orbits within the region, and to reach more distant destinations. Power, thrust, operating life, and radiation resistance will all have to improve for these later uses, so propulsion R&D would be ongoing. An early use for such engines is mining the upper atmosphere for gases, as noted at item 1.3 Extract Materials above. Some of the gases can be used for propellant, which makes the propulsion self-sustaining. With sufficient propellant, mining of orbital debris becomes feasible.New payloads delivered to Low Orbit can also be delivered to their final destinations efficiently. One early category of missions are prospector satellites to observe and return samples from Near Earth Asteroids, to prepare for later mining.

Chemical Propulsion - High-thrust engines, such as conventional chemical rockets, are still an attractive option for some purposes, despite lower efficiency. These include landing on bodies with significant gravity wells, and when velocity change or transit needs to be done quickly, such as passing through the Earth's radiation belts. Which propulsion type to use for what part of a trip will need to consider multiple factors, including the ability to produce propellants locally. R&D for chemical propulsion will include adapting systems to use and store propellants made in orbit, and improving engine operating life.

Spaceport Network - In the long run, large numbers of vehicles changing their orbits by consuming propellant is inefficient and wasteful. Large scale infrastructure which reduces propellant needs would be desirable. We will refer to them as **Spaceports** by analogy to maritime and airports. Their main function is transport of payloads by potential and kinetic energy change. They would also serve as transportation depots, with docking for multiple vehicles, habitation, warehousing, maintenance, fueling, etc. The first concept for such infrastructure was the **Space Elevator**, which dates back to 1895. Unfortunately the Earth's gravity well is too deep for the original idea of a one-piece stationary elevator to work with any known materials. The original idea can work for smaller bodies, and systems with several smaller pieces can do most of the velocity change for Earth using available materials. A network of spaceports can eventually replace much of the propellant used in space, and increase the percentage of payload transported. The R&D work for such a network is placed here because the first spaceport would likely be located in low orbit. Section **4.11 - Space Elevator** looks at some alternative concepts for such a network.

The basic transport function is accomplished by **Momentum exchange** between a payload and the spaceport structure. Depending on direction, the payload gains or loses energy, and the opposite happens to the spaceport. If traffic is balanced, or the spaceport is anchored to a more massive body, its orbit is not affected. Unbalanced orbit changes are corrected by an efficient propulsion method on the spaceport. To the extent this replaces lower efficiency vehicle propulsion, especially when reaching orbit from the Earth's surface, there is a net savings. Various experiments have been done in orbit related to this technology, but much more work is needed. Improvements in other technologies beyond momentum exchange are needed for a complete spaceport network and associated vehicles.

Phase 0H: R&D for Phase 4B High Orbit Development

High Earth Orbits are currently used by a number of remote-controlled satellite types, including communications, scientific, and navigation. They are all delivered from Earth, and local production and habitation don't yet exist. Transport is only that built-in to the satellites when delivered. The High Orbit region is fairly devoid of native materials, but has a high level of solar energy, and is accessible from Earth, the Moon, and Near Earth Asteroids. Between current civilization on Earth, and future locations beyond it, it can serve as a useful production and transport nexus, and later for large-scale habitation. Fully developing this region will require extensive R&D work. Some identified tasks include:

1. High Orbit Production

1.4 Materials Processing - This is the conversion of raw materials to finished supplies or stock materials. In the early stages this can be asteroid materials brought back from Inner Interplanetary orbits to a location near the Moon, such as Earth-Moon Lagrange Point 2 (EML2). EML2 is a Lunar-synchronous location 64,500 km behind the Moon's center. It is low energy to reach from interplanetary orbits, and provides full time sunlight for power. Early products include shielding, propellants, and water. Extensive R&D is needed to identify the best locations and processes. As additional regions in space are developed, raw materials can be supplied from the Moon and farther asteroids, and possibly Low Orbit. Materials coming from Earth will generally be in finished condition, since processing on Earth is less expensive. They would include items like alloying elements for metals, and doping elements for electronics. How finished materials coming from other space regions will be depends on the balance of local processing energy vs transport energy, and what fraction of the ore can be used.

Earlier phases of the program should have developed experience with self-expanding production and remote operations. We assume materials processing begins with finished equipment brought from Earth, then bootstraps further expansion by adding seed factory elements, which use the early supply of processed materials as inputs.Until larger human habitation can be supported locally, it would rely mostly on remote control and automation. Some processing operations may not function well, or at all, in zero gravity, and others will benefit from or work uniquely in the zero gravity and vacuum conditions. So a major research area will be which specific processing flows are to be used under what conditions.

1.7 Assemble Elements

Large Space Structures - Large orbiting structures like the Space Station have been assembled using alignment guides and motorized bolts. For future projects needing large pressure-tight compartments, one option is welding, which is a basic industrial process on Earth. Welding of metals has been achieved by concentrated solar energy **(Romero, 2013)**. Since high-quality solar energy is widely available in orbit, research on using it for welding in space seems worthwhile. For assembling large structures like habitats, where moving the structure would be difficult, one method is to use articulated mirrors to direct a beam of concentrated sunlight at various angles. Another approach to large structures is laying high-strength reinforcing fibers between layers of plasma-sprayed metal. Spools of fiber and metal wire are compact and modular, but the finished structure can be large and seamless.

Phase 0I: R&D for Phase 4C Inner Interplanetary Development

Section **4.7 - Inner Interplanetary Development** describes our concept exploration for Phase 4C of the program. So far we have identified one general and several specific R&D subjects to work on for this phase:

Interplanetary Bootstrapping - Self expanding systems are a general approach used throughout the program. This R&D topic is about how best to grow from early materials extraction from the region, to large-scale finished locations, with a range of production, habitation, transport, and services capacity. A future "space city" is not likely to be built all at once in final form, any more than cities on Earth are. The question is then how to start small and build them in increments.

1. Inner Interplanetary Production

1.3 Extracting Materials

Mining on Earth is very well developed, but extracting materials beyond our planet is still in the early stages of remote sensing and robotic prospecting. Therefore extensive R&D is still needed for this production step.

The general rationale for mining in space, rather than bringing everything from Earth, is based on minimizing total energy use. The Earth's gravity well has a fixed energy cost of 31–62 MJ/kg to climb, depending on orbit. Existing transport methods are inefficient, multiplying the minimum value by approximately 9:1. Destinations beyond Earth orbit require even more energy to reach. The production energy from raw materials to finished products is typically much less than this, in the range of 10–20 MJ/kg. Industrial equipment can normally process many times its own mass, and use many times the energy required to make it during its operating life. So the product/equipment ratio is high. It therefore takes far less total energy to deliver starter production equipment, and make the rest of the equipment and finished products from local materials and energy, than to deliver the all finished products from Earth. Local production includes making propellants for space transport, which makes the delivery of starter equipment to distant locations easier.

Early Mining - Section **4.9 - Orbital Mining** looks at alternatives for supplying raw materials to the Low and High Orbit regions. These would at first come from the Near Earth Asteroid (NEA) group in the Inner Interplanetary region, and be delivered to a processing location near Earth. NEAs are the next easiest to reach materials after the atmosphere and artificial debris in low orbits, and development is likely to proceed outwards from Earth. Remote control of operations from Earth and supplying live crew makes processing closer to Earth is easier at first. High orbits are a convenient meeting point for asteroid, lunar, and Earth source materials, and they also get full-time sunlight. Most of the uses for early products will be in Earth orbits too. So that seems to be the preferred starting point for production.

The mass returned by a mining system and tug from a nearby asteroid to high orbit is on the order of 100 times the equipment mass per trip. The tug's service life is on the order of 6 trips taking 2.5 years each before major component replacement is needed. A typical trip consumes 2.6% of the returned mass in propellant, but certain asteroid types contain up to 20% easily extracted propellant. This propellant consumption assumes the Moon is used for gravity assists in both directions, making the propulsive velocity change less than that to reach Earth escape. So the mining operation can be self-fueling after the first trip. Counting hardware plus initial propellant load, the tug will return about 160 times the starting mass during its life. Mining should drastically reduce operating costs in space, if that mass can efficiently be put to use. Other production equipment will be needed beyond the tugs which bring back the raw materials, but at least the first step has a large positive return.

NEA orbits and compositions are randomly distributed. We prefer to mine the easiest to reach ones at first, when they are in optimal positions and at the best times. Given over 17,000 objects discovered in this group so far, there will be some in such easy to reach orbits. But which ones are best to visit varies with time, because they are in constant motion. 81% of discovered NEAs are estimated to be larger than 30 meters in diameter. They have a mass of at least 18,000 tons and usually much higher. This is too much to move as a unit for early tugs, so R&D is needed on the best ways to collect smaller loads of material from them.

Long Term Mining - As development extends outwards from the Earth Orbit regions, the destination for raw materials will shift to Inner Interplanetary locations, and the source materials will come from the entire region, rather than just the ones easiest to reach from Earth. Depending on the technical details of extraction, processing, later production steps, and final use locations, materials may continue to be transported in the raw state, or production plants brought to the source materials, and more finished products delivered elsewhere or used locally. The size of the asteroid is likely to be a strong factor in this choice. For example, **433 Eros**, a large asteroid in this region, has a mass of 6.69 trillion tons. So it is more likely to be worth setting up local production ton Eros than one of the smaller asteroids which only mass a few kilotons. As more distant orbital regions, and Mercury, Venus, and Mars become developed in later phases, they can also become sources of raw materials to bring to the Inner Interplanetary region.

1.4 - 1.8 Materials Processing & other Production Steps

Materials processing methods, like ore reduction and chemical technology, are also very well developed on Earth. They have seen essentially no use in space beyond some experiments in orbit and prototypes on Earth. So extensive R&D is needed for this and the later production steps. A few uses, like bulk radiation shielding, don't require changing materials from their raw state. But nearly all other materials need some processing to turn raw materials into finished materials inventory. Some materials, like propellants, water, and oxygen, can be used as-is once extracted. Other materials, such as metals and ceramics, need further fabrication into parts, then assembly to make finished products.

Section **4.n - Processing Factory** looks at concepts for the processing part of production. We expect previous designs to have been developed for the Low and High Orbit regions. Additional R&D needed for this phase involves adapting and optimizing the processes for the unique conditions and source material in the region, and brought to it later from other regions. An example future changes is production closer to the Sun, where making use of increased solar flux is desirable. Fabrication and assembly methods may not need changes from previous orbital regions, but this is still to be determined.

3. Inner Interplanetary Transport

3.1 Bulk Cargo Transport - Electric "Space Tugs" are needed to move raw materials from where they naturally occur to where they can be processed, and move finished products and other cargo from place to place. Tugs generally do not need human crews, and are slow but efficient. Electric propulsion has already been developed at smaller scales, but much larger units are needed for this task, and the tugs should be designed for refueling, so they can be used multiple times. We expect smaller tugs to have been developed for the Earth orbit and Lunar regions. So for this region, the main R&D work is on building larger and longer-lived versions.

3.3 Transport People - We want to eventually carry people to open space locations and the major planets and moons in the Inner Interplanetary region. However, radiation is present throughout the area from the Sun and cosmic sources. "Transfer Habitats" are a way to carry people safely and efficiently. These are placed in repeating transfer orbits between bodies or locations, such as between Earth and Mars. Since the habitats don't change orbit once set up, they can have heavy shielding, greenhouses, and processing equipment. The raw materials come mainly from asteroids already in nearby orbits. Local production reduces payload needed from Earth, and gives the crew and passengers something useful to do during the trip. Small vehicles are used to get from the habitats to planetary orbits at each end of the trip. One way to save on high thrust propellant is to use momentum exchange for the small vehicles, making up any needed velocity change on the habitat with electric propulsion. The habitats can grow over time, eventually becoming destinations themselves.

Additional habitats can be set up on the Martian moons as way stations, and eventually other locations. All the locations would eventually become multi-function, combining transport and other purposes. This is possible in space, because unlike Earth, everything is in relative motion. We can make use of that motion for transport while doing other things. This would include producing propellants and other supplies, spacecraft construction and repair, serving as science platforms, and as the nucleus for later permanent colonies. Extensive R&D is needed on how to build and expand such mobile habitats, and the various systems they need, such as food production and environmental recycling.

Phases 0J to 0L: [RESERVED]

The following three sub-phases do not yet have identified R&D tasks. Their sub-phase headings are reserved for later use.

- Phase 0J - R&D for Phase 4D Main Belt & Trojan Development
- Phase 0K - R&D for Phase 4E Outer Interplanetary Development
- Phase 0L - R&D for Phase 4F Scattered, Hills, & Oort Development

Phase 0M: R&D for Phase 5A Lunar Development

Development Sequence – The Moon is physically near the Earth, and visible to anyone who looks at the sky. Despite the obvious destination, the Lunar surface is not the first place we want to start development. This is because the Low and High Orbit regions around the Earth and orbits around the Moon are all easier to reach than the surface. Developing orbital vehicles and supply depots first makes reaching the surface easier, so we start those phases earlier, but then continue them in parallel. Scientific observation of the Moon began as soon as telescopes were available, and local exploration began once rocketry enabled getting close to it. Since 1958, over 100 **Missions** have been attempted to flyby, impact, orbit, land, drive on, return samples from, or use the Moon for gravity assist. Many of these have succeeded, including six landings with people. We expect such science and exploration activity to continue, and provide the basic knowledge needed for useful development. The Lunar region includes two distinct environments. These are the surface and body of the Moon itself, and orbits averaging 35,000 km or less from the Moon's center, where the Moon's gravity is dominant. Each environment requires distinct designs to cope with and make use of the local conditions. Since the Moon orbits our planet, the entire Lunar region is embedded in the larger High Orbit region around the Earth, and moves within it.

Many of the technologies and systems needed for Lunar development are not ready to use today. So significant R&D work will be needed prior to designing and building future Lunar projects. We assign the necessary Lunar R&D work to this phase. Some of that work may be carried out on Earth. Other parts may require using the Low or High Orbit regions, or be directly performed in Lunar orbit or on the surface. The ones which cannot be performed on Earth will require suitable transport and supporting systems such as communications. This in turn may require R&D and projects from earlier phases first be completed. The outputs from the Lunar R&D are then supplied to Phase 5A for their use. An important question is when to start using the Moon in the context of developing other regions, and the level of available technology to do so.

Section **4.12 - Lunar Development** begins the concept exploration process for developing the region. That process includes identifying what R&D is needed for the various locations and projects in the region. We organize and discuss those needs in this section according to major function (production, habitation, transport, and services) and lunar environment (orbit and surface). The list is almost certain to be incomplete and need updates over time. We cannot predict in advance which technologies will work, or prove better than their alternatives. So this information will feed more detailed R&D and program planning on a continuing basis.

1.0 Lunar Production

Bootstrapping Methods – The question of how best to build up industry in the Lunar region as been studied to some degree. For example, **Metzger et. al.** have modeled bootstrapping industry on the Moon, and found 12 tons might be sufficient for a starter set. Under a fairly wide range of assumptions, that starter set could grow to a much larger installation. However, much more study is needed to account for multiple sources of materials, orbital vs surface activities, production methods, and the build-up of infrastructure over time. There is enormous production experience on Earth. However self-bootstrapping from starter sets is still mostly theory on Earth, and production of any kind has never been tried in the Lunar region. Sustained R&D is needed on this subject, both on Earth and for the Lunar region.

1.1 Lunar Orbit Production

Production Locations – The energy from the Lunar surface to orbit is 1.5 MJ/kg. Typical production energies, from raw materials to finished products, are 10–20 MJ/kg on Earth. Production energies are likely to be similar in space. Gathering raw materials from the Lunar surface is fairly low energy, since repeated impacts have pulverized the surface into a **Regolith** of loose rocks and dust. Twice as much sunlight is available in high orbits than the Lunar surface. So the preference appears to be to send materials to orbit for further processing, since it can be completed faster.

High orbit can also be a meeting point for materials from the Moon, asteroids, and Earth. Lunar surface materials are lower density minerals, well mixed from impacts, and low in volatile compounds. This is due to the Moon's high early temperatures and low escape velocity. Asteroids usually did not get heated as much, and their denser components have been exposed by collisions. The Moon's denser materials are trapped deep inside. So available materials from the Moon are different from those found in the major asteroid types. Some materials are rare or absent on both the Moon and asteroids, and are more easily brought from Earth. Using all three sources allows a wider range of processes and products than from the Moon alone. The Moon is likely to be the main material source by mass because of low distance and energy.

1.2 Lunar Surface Production

The preference seems to be for most production to be in orbit. However local production for use locally on the surface will likely make sense, and in some cases so will surface production for delivery elsewhere. Like for orbit, extensive R&D is needed to determine what products and processes will be the most useful, and how to bootstrap from starter sets of equipment. Some candidates include:

Sintered Regolith – **Sintering** forms a solid mass from particles by applying heat or pressure, but not complete melting. Example products are paved landing and building pads , roads, and blocks for structures and shielding. Rocks and dust are widely available on the surface, as is sunlight which can be concentrated. Vacuum conditions make binding the particles easier and reduces losses from heating. It is also a simple process, which can be done robotically. These features make it a good candidate for early production. An alternative or supplement to solar heating are microwaves, which heat from the inside rather than outside.

Direct Extraction of Native Iron – Iron-bearing meteorites have impacted the Lunar surface since Since its origin. From Apollo mission rock samples we know around 0.5% of the surface regolith layer is bits of native iron (**Morris, 1980**). It is generally as small particles formed by exposure reduction, micrometeorite impact, or from the source bedrock. The regolith also has 5–13% iron in the form of mineral oxides, but native iron does not have to be chemically processed, which avoids complexity in early production. Potentially you can extract the native iron fraction with a magnet, then separate it from impurities with a furnace, and sand-cast the result into molds made from the abundant fine particles on the surface. Research is needed into the feasibility of the process, and whether early production of iron is worthwhile relative to more complex chemical reduction. The latter can produce up to 25-30% of the ore mass in the structural metals Al, Fe, Mg, and Ti; 20% Silicon for power, and 40% oxygen for life support. Chemical production can therefore make much better use of a given amount of mined material.

Ceramics and Metals Production – Ceramics, such as bricks and crucibles, and metals of all types, are key elements in any modern production. Extensive R&D is needed in how to extract the desired materials, and convert them to useful products on the Lunar surface and in orbit. Thermal processes are common in both categories, so making solar concentrators and furnaces is an important area of study.

2.0 Lunar Habitation

Low-gravity Effects – This R&D task is to determine the minimum safe levels for people and other living things over extended times. Extensive research has been done on zero gravity, but not on levels between zero and one gee. Low gravity, and even extended bed rest on Earth, are known to have adverse consequences for people. We also do not know the long-term effects of low gravity on plants and animals. Artificial gravity can be supplied by rotation, both in orbit and with surface centrifuges. There may be subtle side effects from artificial over natural gravity. All of this needs to be resolved before long-term Lunar habitats are designed. We expect much of this work will have been done in the earlier Earth Orbit phases, because the same problem occurs there. The natural Lunar surface gravity provides an opportunity for research, while early occupation is limited in stay times.

2.1 Lunar Orbit Habitation

Halo Orbit Station-Keeping – **Halo orbits** are potential production locations, since they are accessible to both asteroid and Lunar material sources and in sunlight nearly 100% of the time. However, they are unstable, so station-keeping is needed to stay in position. Required accelerations are about 120 m/s/year, or 3.8×10^{-6} m/s^2. Solar light pressure from a good reflector amounts to 0.08175 N for a 100×100 m area. This provides the desired acceleration to a 21.5-ton mass. Metallized 7.5-micron **Kapton Film** has a mass of 106.5 kg for this 100×100 m area, or 0.5% of allowed mass. Electric propulsion would consume ~0.25%/year in propellant mass. Kapton films in space have demonstrated long service lives, so they are an example of a potentially a lower mass solution. Since solar panels and furnace reflectors will have significant collection areas, a combination of light pressure on them and placement near the Lagrange point may be sufficient to maintain position. Otherwise additional reflector area can be supplied to control drift. R&D is needed to determine the best station-keeping strategy and design of reflectors when needed.

2.2 Lunar Surface Habitation

Lunar Dust Mitigation – Lunar surface dust is fine and abrasive, and may present other hazards to people and equipment. It can be disturbed by equipment operations, and possibly natural electrostatic effects. Research is needed to determine the best ways to reduce or eliminate dust problems.

3.0 Lunar Transport

This section covers transport systems based in Lunar orbit or on the surface. Systems needed to reach the Lunar region, but based on Earth or Earth orbit are covered under their respective phases. Current transport methods for the Lunar region include chemical rockets and several kinds of electric propulsion. New development is needed for specific lunar systems using these methods, and additional research for newer methods.

3.1 Lunar Orbit Transport

Different types of transport systems are preferred for early, low volume traffic, and for later, high volume traffic, where lower cost becomes more important. The requirements for carrying people are different than for cargo, as are the requirements for orbit-to-orbit and orbit-to-surface transport.

Reusable Landers – Landers are capable of reaching the Lunar surface unassisted, and are suited to early development. The first successful landings were in 1966, have continued since then, and more are expected in the future. However, all such landers to date have been single-use. Future improvements would be to develop a reusable lander, which can refuel in orbit, on the surface, or both.

Orbital Cargo Tugs – Electric tugs are efficient but slow methods to move cargo. They would previously be developed for Earth orbits, but units would can later be based at and refueled in Lunar orbit. Gravitational forces are small in the High and Lunar Orbit regions, so transport between them and to more distant regions is relatively easy.

Lunar Orbit Spaceport– A spaceport is transport infrastructure which makes travel easier, but does not itself travel, much like airports function for airplanes. Such infrastructure makes sense when the frequency and volume of traffic is high. The construction cost can then be distributed over many uses. One transport function is propellant supply. Vehicles then only need to carry propellant for one trip, but can refuel as needed for multiple uses. Another is momentum transfer via structural elements. If traffic is balanced in direction and mass, this requires no net energy. It is faster than electric, but still can use that method to save propellant mass by storing orbital energy in the spaceport's mass. The spaceport can support additional functions beyond the basic transport ones. One example is monitoring and control of uncrewed systems in the Lunar region. The reduced distance relative to Earth enables closer to real-time operation. Another is providing radiation protection and artificial gravity for people. The spaceport would start small and grow over time, as traffic and other functions require. It would also exist as part of a larger spaceport network which enables robust and low-cost travel across the Solar System.

3.2 Lunar Surface Transport

Surface Rovers – Surface vehicles are well developed on Earth, and a number have been operated on the Moon and Mars. However, improvements are needed in load capacity, durability, dust mitigation, and traction. Existing lightweight rover designs are suitable for exploration and site selection. They are probably inadequate for heavier tasks like site preparation and mining. We have no experience yet with maintenance for heavily used machines on the Moon, especially for remote-controlled ones. We also have not unloaded or assembled large vehicles there. These subjects need some research.

Bulk Cargo To Orbit – If much of the processing is to be done in high orbit, an efficient way is needed to deliver bulk raw materials from the surface. Candidates include centrifugal and electromagnetic catapults, and large orbital infrastructure, all of which require significant R&D. The current baseline is chemical rockets, but they have fairly low mass return ratios and are not very energy efficient.

4.0 Lunar Services

:

4.1 Lunar Orbit Services

[TBD]

4.2 Lunar Surface Services

[TBD]

Phase 0N: R&D for Phase 5B Mars Locations

Section **4.14 – Mars Development** explores concepts for developing Mars and the orbital region around it. One approach is to start with a habitat on Phobos. At first can we use local materials from that Moon to support trips to the surface. Since we don't yet know the composition of Phobos, other materials may be needed from nearby asteroids. Since Mars skirts the inner edge of the **Asteroid Belt**, there are many candidates to choose from. At first we produce propellants and crew supplies. Later we can construct spaceport structures to exchange momentum and reach the Martian surface more efficiently.

We already have a number of satellites in orbit about Mars, and landers and rovers exploring the surface. With a propellant supply in orbit, we can start to land more substantial equipment and build up larger facilities on the ground. These can be remote-controlled from orbit until enough habitat capacity is available for full-time crew. Early missions can deliver seed factory components to start local production. With surface propellant production, and later large ground accelerators coupled to orbital spaceports, access to Mars will be much easier in both directions, and large-scale development can proceed.

All of these concepts are preliminary at present, and will likely need extensive R&D before definite project plans can be made, and actually implemented.

Phases 0N to 0S: [RESERVED]

Section **4.N - Later Projects** looks at some ideas for later phases. Since technology changes over time, it is not worthwhile to make too many detailed plans far into the future. Long range concepts can serve as a guide for future research, though. As the time frame gets closer, ideas like these, or new ones developed in the future, can be incorporated into updated program plans. The following five sub-phases do not yet have specific R&D tasks identified yet. Their section headings are reserved for later use:

- Phase 0O - R&D for Phase 5C Venus & Mercury Development
- Phase 0P - R&D for Phase 5D Jupiter System Development
- Phase 0Q - R&D for Phase 5E Outer Gas Giant Development
- Phase 0R - R&D for Phase 6A Interstellar Space Development
- Phase 0S - R&D for Phase 6B Exostellar Development

Section 4.3: Phases 1 to 3 For Earth

The early phases of the program have two main goals. The first is upgrading civilization on Earth. The second is to prepare for and enable the later phase space projects, and the construction of space locations. In this section we summarize the three major phases on Earth that follow the respective Research and Development sub-phases performed on Earth. These are (1) Starter Locations and Network, (2) Distributed and Industrial Locations, and (3) Difficult and Extreme Locations. Since this wikibook is mainly about space systems, we only provide a summary here, and refer to the **Seed Factories** wikibook where they are covered in more detail as design examples. We feel that the idea of self-expanding production is important enough to get a separate discussion, because it can have such widespread applications. However, the early program phases on Earth are connected to the later ones in space, so we provide summaries here for continuity. All parts of the program exist within and interact with the rest of civilization. Therefore we try to use what already exists when possible, and not duplicate existing systems or ones in development. We may propose improvements or alternatives to current systems if there are enough advantages to doing so.

Phase 1 - Starter Locations & Network

Phase 1 in General

Phase 1 builds the first operational seed factory equipment and begins a self-expansion process that continues through the rest of the program. The equipment is located at small sites, such as individual homes, and larger sites with multiple items, such as community-built workshops. Sites which are in easy travel distance from each other, such as within a metropolitan area around a city, are grouped for design purposes into a "Location". Larger distances require more transport time and expense for people and physical items. The emphasis is then more on communications, remote collaboration and work, and transport of high value items rather than bulk goods. The designs for use within a location will therefore be somewhat different than those used across different ones.

The work in Phase 1 then involves two main areas: tasks that happen within a specific site or location, and tasks that involve coordination and collaboration between multiple sites or locations. These are addressed separately in the following headings, but we expect them to happen in parallel.

Phase 1 - Starter Locations

Starter locations use equipment and location plans designed in Phase 0 - Research and Development. This includes program-wide general R&D, and specific R&D from Phase 0B performed for Phase 1. In some cases, the Phase 1 locations will evolve from R&D locations, or inherit prototype and early units of production equipment made in Phase 0. R&D is done in relatively few locations, so in most cases, we expect Phase 1 locations to use newly made or purchased items as a starting point. R&D is a continuing process, so not all designs will be available at once, and initial designs may be improved or replaced over time. The development of a conceptual design for a starter location is described in more detail in **Section 5.0 Personal Production** of the Seed Factories wikibook.

A Personal Production location, once expanded to design capacity, produces mainly home improvement and hobby craft scale products for use by the owner-operators and their immediate community. This is a small enough scale that the starter equipment should be affordable. By directly producing for themselves and their local community, the owners have a reason to participate. The design goal is to meet up to 25% of people's needs and wants on a part-time basis. The owner-operators do not need to give up their current homes and jobs to do this. Even at a small scale, though, producing a variety of products with different materials, equipment, and skills goes beyond what an individual can do in terms of funds, working space, and knowledge. Also, home improvement projects like adding a room often need multiple people at once to carry out. Therefore we design for a community of people working together, rather than for individuals.

Since community participation is likely to accumulate over time, the location begins with a small number of people and some conventional tools for specific categories, such as woodworking and carpentry. The conventional tools are added to over time for additional categories, and starter set machines emerging from R&D are added when feasible. Some self-production and expansion is possible even with conventional tools, such as making a workbench with hand and portable tools. The seed machines raise this capacity by being computer-controlled and general-purpose. The sequence of tools and machines for a given location is not fixed, but driven by community interest and skills. The R&D phase supplies guidance on methods and designs to support community choices. Whatever cannot be done internally by the location is supplied from the surrounding metropolitan area or elsewhere.

Phase 1 - Network

Interest and participation will likely not be limited to a single location in easy travel distance of project members. **Section 6.0 - The MakerNet** of the Seed Factories book develops a conceptual design example for a network of multiple locations. These locations interact physically and electronically to help each other make things. Phase 1 network locations share the same goals as Starter Locations in terms of scale, type of outputs, and meeting up to 25% of needs and wants on a part-time basis.

In addition to making products for current use, a longer-term goal of this phase is to prepare for a future of increasingly smart tools that use automation, robotics, and artificial intelligence to replace conventional jobs. If conventional wage-earning jobs become scarce, people will need another way to meet their needs and wants. One proposal is income transfer programs such as **Basic Income**. This is unsustainable on a large scale, because you run out of places to transfer income from. Government programs are also subject to political uncertainty. We think a better approach is for people to use those same smart tools to meet their needs directly. By developing a network of skills and equipment, and expanding it themselves as the need develops, this can be done affordably. By tapping a wide range of locations, with different skills and resources, communities can better prepare for the future, and increasingly rely on themselves in the later phases. People without the needed skills can participate by buying a share of the production capacity or contributing unskilled work. Existing network members can also provide training for new people, so they can gain the needed skills.

Phase 0B R&D to support such networks includes software and communications for distributing tasks across sites and locations, automatically where possible. It also includes efficient transport between sites and locations, and remote control and assistance between sites using tools like virtual reality with force feedback. That way, even if people and equipment are widely separated, they can still work together. It would also include developing training materials and instructions to build and operate the elements it designs.

Phase 2 - Distributed & Industrial Locations

Phase 2 - Growth Locations in General

By design, all types of seed factories can grow by self-expansion and upgrade. The natural continuation of this process is to increase the scale and intensity of operations beyond the part-time and small scale of Phase 1, towards full self-support and trade beyond the internal community. This will likely not happen all at once, but rather by evolution of phase 1 sites as individuals decide to move past it. Phase 1 equipment can be used to make some of the parts for larger equipment. Starter sets are optimized for flexibility, doing many tasks with few machines, to keep their size small. Full-time use tends to favor more specialized machines optimized for their tasks, because efficiency and time savings are worth more. It also favors designs for higher duty cycles and longer operating lives. Such heavy-duty designs are normally more expensive, but self-production and automation can minimize the increase. Lastly, higher intensity of use favors more automation, since the savings are higher the more times a task is performed. Since the designs are different in this phase, additional R&D is required, and again will not happen all at once, but is supplied incrementally. Additionally, research is needed for the optimum growth paths from Phase 1 systems, and across wider ranges of industries than than those for personal use. Finally, not everyone will choose higher levels of activity, and not all of their equipment upgraded to those levels. So the network of locations will include a mix of Phase 1 and 2 elements.

Phase 2A - Distributed Locations

Phase 2A Distributed Locations include small business and commercial levels of operation, including full-time participation of the people involved. This level can be reached incrementally, through internal growth, but may also be started at new sites directly at that level. The larger scale may require outside funding or partnerships to get started. It may also require dedicated sites due to scale and specialized needs, and from legal constraints on where certain types of activity can happen. This sub-phase is called "Distributed" because the full set of equipment and activities is not likely to be at one physical site, but distributed across multiple sites in a given location. As commercial scale operations continue to grow, some can eventually evolve to the Industrial Locations of Phase 2B, while others remain at the smaller scales, all operating in parallel.

This scale of operation is not limited to production tasks, but can serve the full range of industries. This includes service-type industries that use products, but do not make them. For example, a restaurant requires a building, furniture, and kitchen equipment. All of these must first be manufactured, but the restaurant only produces satisfied diners, not finished products used in further production. So the logical progression of growth is from core machines that are used to make more machines, to machines that make end-use items, to industries that only use end items to operate, but don't make any of their own.

Phase 2B - Industrial Locations for Earth

The final growth phase in moderate locations on Earth is to the industrial scale. The goal for this phase is to serve larger and more widespread markets at the most efficient levels of size and specialization. At this scale, outputs are far beyond the community needs of the owner-operators. As in Phase 2A, this scale can be reached by internal growth from previous scale locations, or by starting new sites at full scale. These can be **Greenfield Sites** (previously unused land} or reconstruction on **Greyfield** or **Brownfield** sites. These are respectively outdated and underused sites which have been developed in the past, or previously used for industrial/commercial use and possibly contaminated. A mix of self-expansion and new construction may also be used.

Starting new sites more strongly favors outside funding, because, for example, half a blast furnace is of no use. To be useful it must be complete, and thus need enough resources at one time to build it. Likewise, a large parcel of industrial land is usually acquired at one time, since later additions from neighboring land may not be possible. Once acquired, there is an incentive to put the land to use right away, otherwise the funds could be more productively applied elsewhere. To the extent the needs for an industrial site exceed surplus production capacity and income, then outside funding for may be required. One way to obtain sufficient funding is to distribute ownership across many people, both project members and outsiders. Outside market forces become more important than internal needs at this scale. Since those forces can't be entirely predicted, distributed ownership across multiple industries reduces risks. Work and equipment can then be redistributed when markets shift, and maintain useful production.

Even with sufficient funding, industrial scale sites require larger work spaces with larger input and output flows. Since land parcel sizes, utility supply, and transport capacity are finite, this tends to limit an industrial site to fewer product types or stages of production. Since the sources of inputs and markets for products are more widespread, transport is relatively more important than earlier phases. **Section 7.0 - Industrial Production** of the Seed Factories book will develop conceptual design examples for this scale. It is currently unfinished.

Most of the output from this phase is intended for use on Earth. A portion will be service, transport, habitation, and production items to reach or be used in space. The industrial locations which supply space-related outputs are still integrated with the rest of civilization. But since they represent specialized industries, and fall into the main subject of this book, we cover them separately in Section 4.4.

Phase 3 - Other Earth Locations

Phase 3 - Other Earth Locations in General

The previous phases were intended for locations on Earth that have moderate environments and surrounding development already in place. This phase continues development on Earth to places that are not as easy. A major goal of this phase is a better quality of life through sustainable development. The Earth's population is growing, and most people want a developed lifestyle. Serving more people at higher levels of development requires more physical resources and energy. Obtaining them from the limited places with moderate conditions is already stressing the capacity and environment in those areas. A way out of this problem is to leverage smart tools (automation, robotics, and AI) in two ways. One is to access more difficult and extreme locations, where additional physical resources can be found. The other is to produce large amounts of renewable energy, which will have less impact on the environment. These are made affordable by bootstrapping self-expanding systems from starter sets. **Section 8.0 Remote & Difficult Locations** of the Seed Factories book will develop conceptual design examples for this phase, but is currently unfinished.

Phase 3A - Difficult Earth Locations

Difficult conditions involve one or more environment or development parameters that are significantly outside the moderate range. Examples include cold tundra, hot deserts, rain forests, high elevations, weak soils, significant depths underwater or underground, or places requiring long travel times to reach. The different conditions will require some modified or new designs to meet them, and therefore some prior R&D work from Phase 0E. Where the locations are undeveloped, with little in the way of local services, then more equipment is needed to start a minimum level of operations, and more transport of needed supplies from elsewhere. Where possible, these are supplied from previous phase locations.

Phase 3B - Extreme Earth Locations

Extreme conditions go further beyond the difficult conditions and reach up to the limits of available technology. Some parts of the Earth, mainly at great depths, are not accessible at all with current technology, and therefore excluded as possible locations. Working in these conditions may need extensive R&D in Phase 0F to advance the technical limits, and to supply new or modified designs to handle the conditions. Experience in difficult conditions will be useful for the more extreme ones, so this phase logically follows 3A, but is in parallel with all previous phases, which continue to operate.

Section 4.4: Phase 2B - Industrial Locations for Space

Most of the development in Phases 1 through 3 is aimed at upgrading civilization on Earth, and expanding it to more difficult environments on our planet. But the Earth is finite, and regions in space have vast amounts of available physical space, raw materials, and energy resources to continue that work. So Phases 4 through 6 are aimed at developing those regions and continuing the upgrade and expansion of civilization throughout the Solar System and beyond. In that process we also want meet the program objectives noted in **Section 4.1**. Several of those objectives benefit Earth, but use space activity to reach them. Conversely, civilization started on Earth, so future development of space has to start here, and will involve continued support and interaction from Earth with the later phases.

This section of the book then addresses the meeting point of Earth and Space - that part of Phase 2B on Earth which supports and interacts with Phases 4 through 6 in space, and in turn helps meet the main program goals and objective. However, we are not starting with a clean slate. Space industry is already global, large-scale, and on-going, and most of the activity happens on Earth. Our concepts must therefore account for these existing projects and activities. We also want to address the following questions:

- What new projects and locations are needed to accomplish our future goals?
- What parts of existing space projects and locations should remain as they are, in parallel with new ones we add?
- Do some of the new projects and locations belong in other phases?
- What industry categories and products are needed in this phase?
- How will the new projects and locations interact with the rest of our program, existing space programs, and the rest of civilization?
- In what sequence should the new projects and locations and their products be built?

Concept Exploration

Our work here is early concept exploration, the first step among many in a project. It is in no sense finished, but rather a starting point for further work. In later sections of the book we are concerned with developing different regions in space. These are unfamiliar to most people, so we begin those sections by describing the regions in general, their environment parameters, and energy and material resources. We expect most of the new development on Earth will be in moderate environments. These are familiar enough that we do not need to provide a general description in this section. Instead, we will note important features as they come up, and when locations in other environments are needed.

Our exploration begins with a survey of the full range of industries on Earth, noting which have current space-related activity, and where there are possible future additions. Since much of the future activity will be based on later needs in space, we have to consider later program phases to determine what will be needed on Earth. We then look at project drivers, including motivations, economics, and technology, to identify which industries can move forward, and when. From past work by others and our own work, we identify specific projects to satisfy the identified needs. We combine all the information into a concept for the phase. This includes a general approach, a list of projects by time and function, how they relate to each other, and to other program phases. In reaching a phase concept we consider different alternatives, develop project details, and make estimates and calculations. They are included as the later parts of this section. Since our work is as yet incomplete, there will be gaps in the discussion. An output from our analysis will be identifying R&D work which will be needed for the respective projects. This is fed back to planning for Phase 0D - R&D for Industrial Locations.

Existing Space Industry

According to the **Satellite Industry Association**, global space industry as of 2016 was US$339 billion/year. These existing projects were previously described in **Section 1.9**. Most parts of existing space activities are actually carried out on Earth, with a relatively small amount of equipment and people in orbit. For example, just one of NASA's **Crawler-Transporters**, which carry rockets at the **Kennedy Space Center**, has a mass of 2700 tons. This is 6.4 times the mass of the entire **International Space Station** (ISS). NASA's total ground equipment and facilities is much larger, and their ~100,000 government and contractor employees dwarfs the six astronauts who occupy the ISS. This high ratio of ground to space activity will continue until we change how we deliver things to orbit and start to exploit resources already in space. Even then, a significant amount of people and equipment will continue to come from Earth, and space-related industries continue to operate here. Our planet will remain an important part of supporting later program phases in space.

New Project Phasing

We will look at changes to existing industries, and new projects that will be needed on Earth, to support later phases of our program in space. To date, production and launch of space equipment has used industrial-scale facilities. Therefore we place the new projects within Phase 2B Industrial Locations. They will make up a subset of all industrial activity in Phase 2B, most of which will be products for use on Earth. The space industry subset strongly interacts with other industries in Phase 2B, and the rest of civilization outside the program. For example, a rocket launch site will typically obtain concrete, steel, and electricity from outside industries, rather than producing them locally. Such locations will mostly be in moderate environments, but some may end up in difficult or extreme ones. When those are identified they will be assigned to Phase 3 as needed. Some of the activity may be small enough in scale to fall into Phase 2A Distributed Locations, and some of the long-term projects have an orbital component. They will be mentioned here as they come up, and assigned to their respective phases later.

Industry Survey

Phase 2B as a whole covers all types of industrial-scale projects in moderate environments on Earth. In this section we are concerned with the subset which are needed to support later parts of the program in space. Where existing and expected development are already sufficient, we note that, but do not go into much detail. Where additional or unique projects will be needed at some point in time, we try to note what they are. Our list of industry categories is drawn from the latest version of the **North American Industry Classification System (NAICS)**, and we adopt their numbering system. This allows for easier comparison to other data about industry on Earth.

11 - Agriculture: People in space, and those working on space-related projects on Earth, need food to live. Agriculture is already well-developed on Earth, so supply to projects on the ground should be sufficient. For people working and living in space, they will need packaged and storable food to the extent they cannot grow it locally. Once local production is set up in space, there are likely to be food items which are not practical to make locally. There will also likely be a need for modified organisms for the space environment, agricultural equipment, fertilizers, and trace elements which cannot be produced or found locally in space. To the extent they cannot be provided from some region in space, they would need to come from Earth.

21 - Mining: Extraction and processing of raw materials is also well-developed on Earth. Supply of such materials for space projects on Earth is expected to be sufficient. An example is crushed stone and steel to build a new rocket launch site. Hardware to be made on Earth, and delivered to and used in space, sometimes needs specialty materials. Once mining and production is established in space, there are likely to be rare materials and components which are still better supplied from Earth. If those exceed existing Earth industry or need custom design, they may need new projects to support them.

22 - Utilities: Moderate locations on Earth either already have sufficient utilities, or it is straightforward to add them to support new industrial locations.

23 - Construction:

31-33 - Manufacturing:

42 - Wholesale Trade:

44-45 - Retail Trade:

48-49 - Transportation and Warehousing:

51 - Information:

52 - Finance and Insurance:

53 - Real Estate, Rental, and Leasing:

54 - Professional, Scientific, and Technical:

55-56 - Management and Organizational Support:

61 - Education:

62 - Health and Social Services:

71 - Arts, Entertainment, and Recreation:

72 - Accommodations and Food:

81 - Other Services:

92 - Public Administration:

Project Drivers

Motivations

Economics

Technology

Placement

Development Projects

We first describe our general approach to Phase 2B projects for space, then organize the ones we have identified by time frame and major function. Since reaching space is necessary to carry out the later phases, a large portion of the projects will fall to the transport function.

General Approach

Extensive space industries and projects already exist on Earth, and are likely to continue for the forseeable future. Where these existing projects are useful, we would keep them as they are. As changes and new projects are needed, they would be introduced gradually over time. Current production and operating locations would be re-used to the extent possible to lower costs. We organize projects for this phase by time into four groups. The ones farther in time typically depend on earlier ones to get started, require more R&D to get ready for them, or await markets developing to the point they are needed. Later projects will also benefit from better technology and capabilities from other parts of our program, and from the rest of civilization. The four sets by time are:

- **Current** - those already operating, or have started detail design or are in later stages of development.
- **Near-term** - those that are planned to reach detailed design within 10 years, and have significant funding sources available.
- **Mid-term** - those which can reasonably start detailed design in 10-30 years, and may or may not have funding sources.
- **Long-term** - those which will likely need more than 30 years to reach detailed design.

Large-scale non-space industry also exists on Earth, unlike space, which starts out undeveloped. For new projects, this make the incentives to bootstrap from starter sets and use local resources less on Earth than later phases in space. The seed factory approach would have been developed in earlier phases, and used in parallel for non-space industry in Phase 2B. So we include it as one of the tools of modern engineering and apply it to space industries when it is useful. The remainder of the industrial-scale projects for space are built and operated conventionally, by importing equipment and materials as-needed from other industries, either within the program or from outside it.

Industrial Production for Space

Transport to space requires supplying cargo containers at a minimum, but usually more complex vehicles. Making these falls to the production function. Transport also requires some type of ground facilities. At a minimum the ground facilities support vehicle operations, but in some cases do much of the work of accelerating payloads to orbit. Building the ground facilities is also assigned to the production function. Operating the vehicles and ground facilities falls to the transport function. Whatever type of cargo is going to space must also be produced, or at least acquired. When produced internally they are also assigned to this function.

Industrial Habitation for Space

This function in general includes large scale construction which is intended to be occupied by people. Examples include large office buildings, hotels, residential towers, retail complexes, and entertainment venues. Space for people to live, outside of their work for the program, is generally well-developed on Earth. So our program does not have to provide it, except to note where significant additions are needed. Office, laboratory, and other space to accommodate people working within the program are included in this function. In the case where people are living in a remote location, such as a floating ocean launch site, their living space would be provided by the program, and therefore included.

Industrial Transport for Space

Phases 4-6 include future Orbital, Planetary System, and Interstellar Development. For any of that to happen, we first have to deliver people and equipment to space. The lowest stable orbits at 160 km altitude require 30.48 MJ/kg kinetic and 1.53 MJ/kg potential energy to reach. Those are ideal values, with current rockets consuming about 285 MJ/kg. This is because hemical rocket propellants don't contain enough energy relative to what is needed to reach Earth orbit. So existing rockets need a very large ratio of propellant to payload mass, yielding about 11% overall efficiency. Useful low-orbit payloads average 1,500 kg in mass, therefore needing 427.5 GJ per launch. A minimum reasonable launch rate is 6 times a year, so the total energy use is over 50 times annual average US electric consumption/capita. Beyond energy needs, transport systems need hardware and operations support, making such projects industrial scale activity, and therefore part of Phase 2B.

Before 2015, the vehicle hardware was mostly used once and thrown away. To enable a useful payload mass while carrying so much propellant, the hardware had to be high performance and light weight. It was also produced in fairly small numbers. This made the hardware expensive, and throwing it away led to very high costs to reach space. Some post-2015 programs are developing reusable rockets. This will partly solve the cost problem, but the underlying inefficiency of chemical rockets remains. Our approach for the mid- and long-term is to replace part or all of the transport to Earth orbit with methods that are inherently more efficient. Alternately or in addition to this they would use equipment produced in larger quantities, or with less extreme performance needs, leading to lower costs.

The transport function in general includes large scale delivery of energy, discrete and bulk cargo, fluids and gases, people, and data. Current transport to orbit uses the same vehicle for cheap bulk items, like propellants, as for people and high-value equipment. Safety, reliability, and other features are driven by the needs of the latter payload types. Those features are applied to the bulk payloads too, even if they don't need them. On Earth, we use different kinds of transportation depending on what is being carried. So another part of our approach is to use different transport methods suited to their respective payloads, when that makes sense.

Industrial Transport Alternatives

Phases 4-6 cover a long period of time, and have a wide range of potential needs for transport from Earth. So the space transport portion of Phase 2B will also have to cover the same range times and needs. Our analysis will therefore look at a wide range of potential alternatives. The baseline option is to stay with currently existing launch systems, and those already under development or planned. One alternative is to keep the baseline systems, but add items like more launch pads and vehicles to increase capacity. Another is to explore new transport systems to be developed within our program. For current systems, we include ones already operating, and those which have entered detailed design and production. For planned we include those expected to start detailed design in the near-term (within 10 years), and have significant funding sources available.

For new systems, one alternative is another conventional launch system, of the types already operating, but sized to meet our Phase 4-6 needs. Such a launch system may or may not have outside funding, but we assume using advanced production, of the type developed for our program, to lower costs. More advanced alternatives can be grouped into those that augment or supplement chemical rockets, but still use rockets as the primary method to reach orbit, and those which substantially or completely replace chemical rockets. The latter may depend partly on orbital systems which would be part of Phase 4A.

Reasons to explore alternatives include not enough capacity in mass to orbit, or costs too high to make projects in later phases feasible. Additional reasons include improving system efficiency and lowering cost for business reasons. The development cost and complexity for new systems have to be weighed against the performance and payload gains they generate. Where new or more advanced methods are used they typically add substantial R&D time and cost. There is also a technical risk of the methods not working as intended, or less well than desired. The extra time, costs, and risks must then be weighed against the limitations of current and near-term systems to determine a preferred set of concepts. Our exploration of these alternatives begins with identifying what they are. In the concept details section below we compile later phase needs in terms of time and traffic. Finally, for each alternative, we attempt to estimate performance, costs, and risks.

Identifying the Alternatives

Conventional Alternatives:

Current and near-term launch systems include a large variety of **Multi-Stage Rockets** and a few **Air-Launched** ones, where a carrier airplane takes the rocket above most of the atmosphere before ignition. Information about these systems can generally be found in their **User's Manuals** when they are in operation or later stages of development. For those in earlier stages, information can be found in public sources or by contacting the projects directly. For our current purpose, general information is sufficient. For new conventional systems we can design another pure rocket or subsonic air-launch system similar to existing ones, but sized to meet the expected traffic.

Augmented Rocket Alternatives:

Augmented rockets use technology beyond what is currently used for launch to orbit. They still use chemical rockets for over 80% of the velocity needed. Since payload is non-linear with rocket velocity, a 20% reduction in the rocket portion can result in 50-80% increase in payload. Examples in this group include ejector rockets, high Mach carrier aircraft, aerostats, vertical jet boosters, and low-G gas accelerators. A combination of methods may be used to reach the 20% level.

Rocket Replacement Alternatives:

Part 2 of this book provides an extensive list of space transport methods. We consider the subset which can replace 20-100% of the velocity from the Earth's surface to low orbit from conventional rocket stages. To do this they must overcome the Earth's gravity while contributing to orbital altitude and velocity. They must also operate safely in the atmosphere, and be feasible in the time frame of the later phases, both in technology readiness and cost. Multiple transport methods may be used to replace a higher percentage of the total velocity, including some of the augmented group. Examples include high Mach combined cycle air/rocket engines, hypervelocity gas or electromagnetic accelerators, and orbital spaceports with an elevator system. The last of these would be in low orbit and belong to Phase 4A, but can supply part of the velocity needed.

Traffic and Schedules

Which systems are suitable for later phases will depend on **Traffic Models** for the phases. These are launch needs in terms of payload sizes, mass, and quantity by year. Our models would be drawn from the needs of later program phases in space, and necessarily become more uncertain in later years. The long-term portion will mainly identify what early transport R&D should be invested in, so that they will be ready when the need arises. In developing a traffic model, larger payloads will need dedicated launches to their desired destinations. Smaller ones can fly as secondary cargo on an "as available" basis, when the main payload leaves some unused capacity. Transport capacity to fill the model's needs, for both existing and new systems, can be tabulated in terms of payload and number of launches per year.

Performance, Cost, and Risk Estimates

Full-scale design of any large-scale system has to consider many factors. At the concept exploration stage, the most important are performance, cost, and risk. For transport to orbit, performance includes physical size, mass, destination orbit, and launch schedule. The schedule includes necessary time for R&D, design, and production for facilities and vehicles. Cost includes R&D, design, and production costs, plus ongoing operations costs. Technical risks are the uncertainties that the system will work at all, less well than desired, or suffer failures. Cost risks include availability of funding, uncertainties in development and operations costs, and market needs for the system. The more advanced and farther in the future a given alternative is, the less accurate our estimates become. That includes technology improvements that will happen outside our program. Even uncertain estimates are useful to identify R&D investments that have high potential gains. Some high-risk/high-payoff investments will not work out. This is acceptable if enough of them do to justify the overall R&D effort.

Industrial Services for Space

[TBD]

Program Integration

[TBD]

Transport Details

Conventional Alternatives

Current & Near-Term Systems

Current Conventional Rockets:

The following is a subset of significant conventional rockets which have flown recently as of 2017, and multiple times previously, which we consider current. Payload masses are to Low Earth Orbit unless otherwise noted. Variant rocket configurations yield different payload masses.

* **Antares** (US, Orbital ATK) - 6,500 kg
* **Ariane 5** (ESA, Arianespace/Airbus) - 16-20,000 kg
* **Atlas V** (US, United Launch Alliance) - 9-20,500 kg
* **Delta IV** (US, United Launch Alliance) - 13-28,899 kg

Near-Term Conventional Rockets:

This includes rockets expected to reach detailed design by 2027.

* **BFR** (US, SpaceX) - 150,000 kg

Current Air-Launched Rockets:

Near-Term Air-Launched Rockets:

[To Merge]

As of 2012, the following launchers are specifically designed to transport humans, which requires a pressurized environment and other design features. Other existing and under-development launchers can deliver cargo, and some of those could be adapted to carrying humans.

Soyuz

Shenzhou

Crew Transportation System

This is a NASA funded project with multiple private sector contracts to develop components and ultimately a functioning transport system. As of April 2012, proposals for the next stage of development were being reviewed by NASA.

Space Launch System

Stratolaunch

Skylon

A project from Reaction Engines Limited (REL) with Alan Bond directing the efforts. The Skylon spaceplane is designed as a single-stage-to-orbit craft, that can take off and land like a normal airplane. The engine is built around a hybrid approach, it functions both as a normal air-breathing engine (jet) and a rocket (in the high atmosphere). This setup is intended to reduce the amount of oxidizer propellant required to send cargo into space as to cut costs.

An European Space Agency (ESA) design evaluation commissioned by the UK Space Agency (UKSA) and concluded in May 2011 stating that "ESA has not identified any critical topics that would prevent a successful development of the engine,".

As of April 2012 the funding of the project was mostly from private investors 85%, and funding is still being sought to complete the project. The Reaction Engines Ltd **Skylon Users' Manual** (Rev 1, Sep 2009) gives extensive detail about the vehicle and it's engine.

New Conventional Rocket

Conventional rocket design has been done many times in the past, and is well understood. We refer you to any of a number of texts and references on the subject, such as **Rocket Propulsion Elements** (9th ed., Sutton & Biblarz, 2017), for more detail. We give an example of a small multi-stage rocket to provide a general overview of the design process. That process starts with some initial assumptions, from which we can make an estimate of the vehicle size. We then progressively add more detail and do more accurate estimates. This will replace our initial estimate with a series of better ones, and possibly force revising the assumptions. A complete preliminary design considers all the major components and is at the point where you would start the detailed design and final drawings. We will not carry it that far, but want to show enough of the process to show how it is started.

Design Assumptions

- Payload: 20 kg to 250 km circular orbit - This is very small for a practical system, but the same formulas work at any size. We need to specify an orbit to calculate the mission velocities.
- G-Limit: 10 gravities or 100 m/s^2 - This limits accelerations and structural loads on the payload. Larger payloads are typically limited to 6 g's, but ones this small can withstand higher acceleration without much penalty
- Exhaust Velocity: 3300 m/s in vacuum - This is typical of a moderate performance engine using Methane/Oxygen propellant mix.
- All stages are re-used for cost reasons. Hardware mass fractions are assumed to be 14, 15, and 18% for the first to third stages. The upper stages would have higher fractions due to smaller size and increasing heat shielding.
- Launch Site: Equator at 4600 m altitude - This is at Cayambe, Ecuador to take the most advantage of the Earth's rotation and highest starting altitude to reduce drag and mission velocity

Preliminary Estimate

Conventional rockets are sized by the **Rocket Equation**, which determines propellant mass ratios. A preliminary estimate of the velocity required can be made from experience. A second estimate will use a trajectory simulation that calculates fuel use, thrust, drag, and acceleration in small time steps.

The ideal velocity to reach a 250 km orbit, neglecting losses, is found from the total energy of that orbit, which is the sum of kinetic and potential energy. Velocity is 7756 m/s, with an energy of 30.08 MJ/kg, and potential energy is 2.375 MJ/kg. The sum implies a velocity of 8,056 m/s. The various real losses may be estimated at 900 m/s based on experience, giving a total ideal velocity of 8956 m/s. Rotation of the Earth at the Equator is 465 m/s, thus the rocket has to produce a net velocity of 8,491 m/s. If we divide it equally into 3 stages, this gives 2830 m/s per stage. Mass estimates are calculated from top to bottom as follows:

- Payload = 20 kg
- Stage 3 final mass /initial mass = 42.4% - From rocket equation
- Stage 3 hardware fraction = 18% - of entire stage including payload
- Stage 3 initial mass = 20 kg / payload fraction = 20 kg / (final mass - hardware) = 81.9 kg
- Stage 2 m(f)/m(i) = 42.4%
- Stage 2 hardware = 15% x (100-42.4%) = 8.64% - of 2nd stage fuel only
- Stage 2 initial mass = 81.9 kg / (42.4% - 8.64%) = 242.5 kg
- Stage 1 m(f)/m(i) = 42.4%
- Stage 1 hardware = 14% x (100-42.4%) = 8.06% - of 1st stage fuel only
- Stage 1 initial mass = 242.5 kg / (42.4% - 8.06%) = 706 kg

Second Size Estimate

To make a second estimate we need some details of the rocket thrust and drag, and therefore it's size and shape. We assume Oxygen/Methane fuel at 3.6:1 **mixture ratio** by mass. The chemistry of CH_4 +

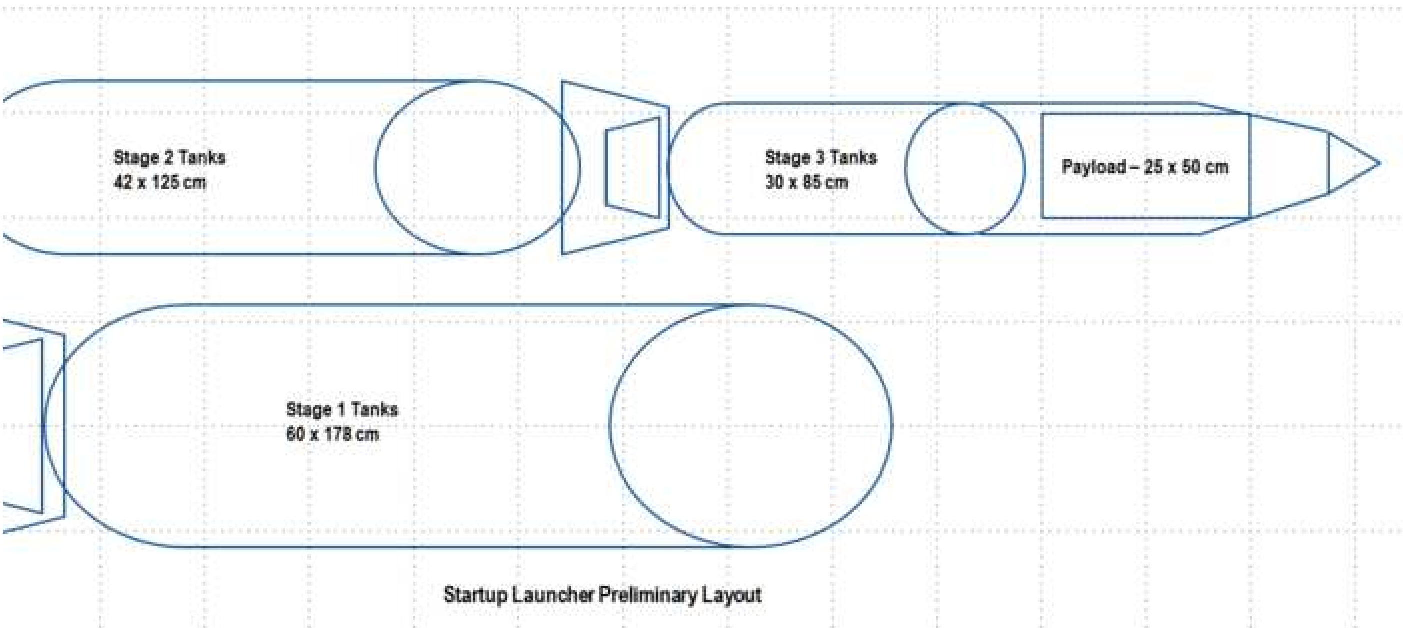
$2O_2 = CO2 + 2H_2O$ has a theoretical mass ratio of 4 Oxygen : 1 Methane. By using slightly less Oxygen some of the Methane is left unburned, leaving CO or H_2 in the exhaust. This lowers the average molecular weight and increases the exhaust velocity. It also ensures the combustion is not Oxygen rich, which would tend to react with surrounding materials.

- **Tank Sizing:**

From our preliminary masses above, we can determine tank sizes from the density of the respective fuels: Oxygen = 1140 kg/m^3 and Methane = 423 kg/m^3:

- From above, all stages have a final mass of 42.4% of initial mass, therefore burn 57.6% of initial mass in fuel Therefore fuel masses are 406.7, 139.7, and 47.2 kg.
- With a mixture ratio of 3.6:1, the Methane component is 1/4.6 = 21.74% by mass, and Oxygen is the remainder. Thus the Methane mass by stage is 88.4, 30.4, and 10.25 kg, and the Oxygen mass by stage is 318.3, 109.3, and 36.95 kg.
- From the densities we can calculate the respective tank volumes. Allowing 3% extra volume so that there is some pressurizing gas at the top of the tank and fuel margin, we obtain first stage tank volumes of 215 and 288 liters, second stage of 74 and 98.75 liters, and third stage of 24.95 and 33.35 liters for Methane and Oxygen respectively.
- Rocket stage tanks can share a common wall between fuel and oxidizer if they are fully sealed, and usually use an ellipsoidal dome with a 70% height ratio to minimize structural mass. We assume the payload has a density of 1 kg/liter, and thus requires 20 liters volume. For aerodynamic and structural reasons we want to keep the total vehicle height at 10 times the base diameter or less. Each combined stage tank can be modeled as two ellipsoidal domes plus a cylinder. Applying some geometry results in tank diameters of 60, 42, and 30 cm.

- **Drag Coefficient:**



From the tank sizes we can do a preliminary layout of the vehicle. We have to include a forward payload fairing and aft engine sections for each stage to get the total height of the vehicle. For this design we assume an aerospike type engine with platelet injectors for each stage, which gives a total vehicle height of about 6 meters. The layout shown here is not intended as a design drawing. It is a schematic sketch to estimate size and shape of the cylinder and cone sections, from which the drag can be estimated. The layout grid lines are at 25 cm spacing.

At our assumed launch altitude of 4600 m, air density is 0.769 kg/m^3, velocity averages 120 m/s in the subsonic region, the rocket length is 6 meters, and the reference viscosity of air is 18.27 x 10^{-6} Pa-s. Therefore the Reynolds number, Re, averages 30.3 million, but it will change with altitude and velocity. From reference data as a function of velocity and Re, the skin friction coefficient, C_{sf}, will vary from about 0.0032 at low velocity, to 0.00245 at 120 m/s, to 0.00215 at 240 m/s. This is adjusted by a correction factor based on the shape of the rocket, which in this case is 1.085, and the wetted area to cross section ratio, which is ~8.3/0.283 = 29.3. So the total drag coefficient will vary from 0.102 to 0.078 to 0.068 at the given speeds, based on cross section area. Drag coefficients at transonic and supersonic velocities are different, but found through similar steps.

If the vehicle had a base area exclusive of the nozzle, we would need to add base drag. In the case of a functioning rocket, the exhaust fills the base and there is no low pressure area to create a net force by pressure difference relative to the front. If the vehicle flies other than directly pointing in the direction of motion, there will be an additional component of drag due to lift, but for this estimate we assume a zero-lift trajectory for simplicity

- **Trajectory:**

The launch trajectory cannot be determined by a simple formula or graph, because the thrust, drag, and mass of the vehicle are all varying continuously. Therefore a simulation must be done in small time steps so that the above parameters are nearly constant within each step. If the average values within each step are close to correct, then the total trajectory will be nearly correct. This is too many calculations to do by hand, so a computer program or spreadsheet is used. The simulation takes as inputs variable vehicle masses and a **Trajectory Profile**, which is how the vehicle tilts vs time and varies thrust or does staging. The inputs are varied until the desired payload mass and orbit is reached. Modern trajectory simulations will vary the inputs automatically to find an optimal trajectory profile.

- **Reference Concept:**

With a known trajectory profile and propellant masses, the major dimensions of the stages can be determined, and a reference concept for the overall vehicle prepared. Preliminary design can proceed from this point to include layout of the engines and other major components, and their masses. From the vehicle design, preliminary work on the supporting ground systems (launch pad, handling equipment, storage tanks, etc.) can be started. Since we assumed a specific launch site, a site plan can be developed using the actual geography

Augmented Rocket Alternatives

The augmented rocket category still uses chemical rocket stages for at least 80% of the velocity change to reach orbit, but different or higher performance methods than sub-sonic air-launch used currently or in the near-term. These alternatives are not in any particular order. A considerable amount of work will be needed before we have reliable estimates for these systems. So we cannot yet choose among these and the other alternatives. For now we provide whatever details and calculations we have available.

Ejector Rocket

This is a low grade augmentation by entraining air flow with the rocket exhaust. It increases thrust in the first stage by increasing mass flow.

Carrier Aircraft

Current carriers are limited to subsonic speeds. More advanced ones can potentially reach about Mach 5 using ramjets.

Aerostats

Lighter-than-air platforms can reach higher altitudes than winged aircraft, providing a better starting point for launch.

Jet Boost

Rather than using a carrier aircraft, this approach uses high thrust/weight jet engines as a first stage for vertical launch and landing.

The **Stratolaunch** system currently in development uses a subsonic carrier aircraft. The Jet boost launcher uses military fighter engines to reach supersonic speeds and higher altitudes. Both systems share the idea of using air-breathing engines for the early part of the flight, which are 4-20 times as efficient as rocket engines. They also avoid using rocket engines in the least efficient part of their operating range: going vertically, which causes gravity loss, and through dense air where you have drag and engine pressure loss. Jet boost dispenses with most of the carrier aircraft by using vertical launch and landing. Using wings allows getting more mass off the ground, but they also limit operating altitude. Less hardware to develop should lower the development cost. The engines are mounted to a **Booster Ring**, which in turn carries the rocket stage. The booster ring lifts the rocket to around 15 km altitude and 480 m/s (Mach 1.6) velocity. The rocket ignites and continues it's flight from there, while the booster ring returns to a vertical landing at the launch site.

Early Version

For human transport, the minimum capacity is 1 person. Extrapolating from the SpaceX Dragon capsule mass, which carries up to 7 people, we estimate total mass to orbit as 1,500 kg, of which 750 kg is passenger and life support, or uncrewed low g cargo. In an early version the Skyhook would not be present and the launcher is used to deliver the first components for orbital assembly. Air-breathing boosters function better with more air, so unlike an all-rocket system, they prefer to launch at low altitude. We assume a sea-level equatorial launch site. For a 200 km altitude circular orbit a delta V of 7,900 m/s is required from 15 km, including potential and kinetic energy. The Earth's rotation contributes 465 m/s, and gravity, drag, and pressure losses are assumed to be 200 m/s from that starting altitude. Therefore the net velocity for the rocket stages is 7635 m/s.

We assume a re-used two stage chemical rocket with exhaust velocity of 3350 m/s, similar to the SpaceX Merlin 1C extended nozzle engine. Since ignition of the rocket is at altitude, we optimize it for vacuum thrust, which is effectively the operating condition after the first 20 seconds of operation. We increase the Falcon inert mass from 6.5% of stage mass to 11% of stage mass to account for heat shield and other stage recovery hardware so it can be used again. Each stage is assigned 50% of the required velocity, so the calculations are as follows:

- Stage 2 delta-V = 3817 m/s.Mass ratio = 3.125, so final mass = 32% of start massStage inert = 11% x 68% of start mass fuel consumed = 7.5% of start mass. Thus payload = 24.5% of start mass, and also equal to 1500 kg from above. Therefore Stage 2 start mass = 6,122 kg.
- Stage 1 delta-V = 3818 m/s.Mass ratio = 3.126, so final mass = 32% of start massStage inert = 11% x 68% of start mass = 7.5% of start mass. Thus Stage 2 + Payload (what the first stage has to carry) = 24.5% of start mass, and also = 6,122 kg, thusStart mass = 24,989 kg, which we round up to 25,000 kg.
- A modern fighter engine such as the PW F-135 generates 191 kN thrust on full afterburner at sea level. For performance reasons, we want to take off at 2.0 gravities, thus the allowed mass is 9.74 tons per engine. The engine itself (1700 kg) , fuel (450 kg), and booster ring hardware (590 kg) has an estimated mass of 2.74 tons . Thus each engine can lift 7 tons of rocket stages and payload, and we need 4 engines for the 25 ton rocket with some margin.

The net payload to orbit of 3% of the rocket initial mass is not remarkable, but the ability to recover and use all the stages repeatedly is. Liftoff mass of the booster ring + rocket is 36 tons, about an order of magnitude smaller than the Falcon 9 vehicle + Dragon capsule, and it should therefore be proportionally less expensive to develop. If not too much low-g cargo needs to be delivered to orbit, or if other launch systems reach comparable operating costs, then this system may not be justified. Buying launch capacity from someone else would be less total cost.

Advanced Version

For an advanced version, we assume the Skyhook is in place and reduces the required velocity rocket to 4,810 m/s. For this version we assume a single rocket stage, and keep other values as above. The mass ratio is then 4.2, leaving 23.8% of start mass after rocket burn. Net cargo mass is 12.8% of rocket initial mass. With a 20 ton rocket stage, that provides 2.5 tons cargo to the Skyhook, or about 3 human passengers. If larger payloads are desired, then the booster ring would need more than 3 jet engines. A reasonable limit would be 8 jet engines, which can lift up to 56 tons of rocket stage, and deliver 7.15 tons of cargo.

Low-G Gas Accelerator

Low pressure gas in a pipe, typically on a mountain, provides the initial velocity for a rocket. For people and complex equipment the acceleration is limited, allowing up to about Mach 5 at the end of the pipe, after which rocket stages take over.

Gas Accelerated Ramjet

Ramjets are mechanically simple compared to turbine type jet engines, so potentially low cost. The drawback is they do not function at low velocity, so for this alternative we assume a low acceleration gun is used to reach sufficient velocity for the ramjet to operate. At higher velocity, ramjets lose performance, so the vehicle will use rocket power to finish the mission.

The gun location is assumed to be on a mountain slope with a barrel length of 6 km, and the ends at 3200 and 4200 m elevation, such as the SW slope of Cayambe, Ecuador. Acceleration is limited to 6 g's (60 m/s^2) for human passengers, so the muzzle velocity is 850 m/s (Mach 2.8). An uncomplicated ramjet will operate roughly over a 2:1 velocity range. Beyond that requires more compensation in inlet shape and combustion conditions, so we assume the maximum velocity will be 1700 m/s. Average equivalent exhaust velocity is about 14 km/s over this range, using hydrocarbon fuel. We will assume single stage to orbit and do calculations purely on theoretical performance for now.

Single Passenger Scaling

For a single passenger minimal system, we again assume a 1500 kg capsule with 750 kg of delivered human + life support, or low g cargo. Calculations are as follows:

- Rocket mass: 12,500 kg -The rocket stage needs to supply 5,900 m/s net, which implies a mass ratio of 5.88, or 17% final mass. With 11% harware mass, we end up with 6% payload. Our initial rocket mass is therefore payload in kg/payload in percent = 12,500 kg, or about 1/3 lighter than the Jet Boost concept.
- Ramjet thrust: 400 kN - At an average climb rate of 210 m/s, we want the ramjet to gain 850 m/s velocity over 40 seconds, or a little over 20 m/s^2. Therefore the ramjet thrust needs to be 250 kN for acceleration. Drag is roughly estimated at 150 kN, so total engine thrust is estimated at 400 kN (90,000 lb) A very rough estimate of engine size would be 1.0 m^2 in area. Since this is less than human passenger capsule size (1.6 m seated), the passenger size will govern barrel diameter
- Ramjet mass: 3950 kg - Ramjet Thrust to engine mass ratio averages about 20:1, thus the engine will have a mass of around 2000 kg. Fuel required is about 1150 kg, and remaining ramjet related parts about 800 kg. So total ramjet stage would be 3,950 kg.
- Total mass: 16,500 kg - By adding the rocket and ramjet stages, or about 40% less than the jet boost. It should be emphasized that these are preliminary calculations.
- Gun pressure: 500 kPa - A 1.6 m barrel accelerating 16,500 kg at 60 m/s^2 requires a total force of 990 kN. Dividing by the barrel area gives a pressure of 492 kPa (71 psi). This is not expected to be a difficult challenge from a technical standpoint. More of a challenge will be installing 6 km of pipe on a mountain.

Small Prototype Scaling

To build a small scale demonstrator for this concept, let us assume a payload of 20 kg to orbit, with a higher allowed acceleration of 10 g's, and a two stage rocket. The higher acceleration allows us to reach 900 m/s over a shorter barrel length of 4 km, and the ramjet function up to 900 m/s. The net velocity for the rocket stages is then 5,835 m/s, or 2918 m/s each. For a smaller size we assume slightly lower exhaust velocity (3300 m/s) and higher hardware fraction (15%). Mass ratio for each stage is 2.42. Weights are calculated as follows:

- Stage 2 final mass = 1/mass fraction = 41.3%
- Stage 2 payload mass = final mass - hardware = 41.3% - 15% = 26.3% = 20 kg (by assumption)

- Stage 2 initial mass = 20 kg / 26.3% = 75 kg
- Stage 1 final mass = 41.3% (same velocity as 2nd stage)
- Stage 1 fuel used = 1 - final mass = 58.7 %
- Stage 1 hardware weight = 15% x fuel used = 8.8%
- Total Stage 1 = fuel + hardware = 67.5%
- Stage 2 then = 32.5% of launch weight.
- Total mass = Stage 2 / 32.5% = 231 kg

At 20 m/s^2 acceleration, the ramjet needs to provide about 5000 N thrust (1100 lb), which only requires roughly 1/80 square meters engine area. The rocket stages can be represented by a cone 0.5 meters in diameter and 3.5 meters tall with a density of 1, so the engine is small relative to the rocket stage diameter. Ramjet mass would be around 25 kg, and fuel used about 15 kg. Total launch mass would then be 271 kg. Allow 29 kg for carrier/sabot to fit the barrel, and we have an accelerated mass of 300 kg. At 100 m/s^2, the acceleration force then is 30 kN, and the required pressure is 152 kPa (22 psi).

Gas-Accelerated Rocket

You can launch people and delicate cargo with a gas pressure type accelerator if you lower the g forces sufficiently. That forces the barrel length to be as long as possible, so we need to look at geography to select a location. Two good locations present themselves, although others may be possible.

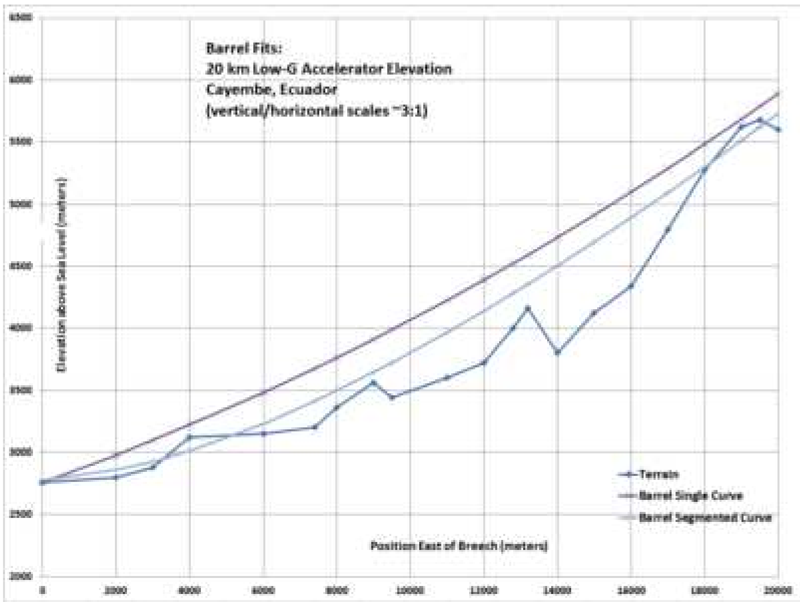
Island of Hawaii

Hawaii is the best location on Earth as far as a large constant slope mountain, requiring minimal grading and support for the barrel, and so lower construction cost. An equatorial site would be preferred to meet up with the Skyhook, but let us first look at Hawaii. It is a shield volcano and cooling lava flows at a constant slope. Therefore you have a nearly perfect ramp on the west side of the island pointing up to the east to build on about 22 km long. You could get as much as 100 km if you extend down into the ocean or add support towers on the eastern slope, but that would be more expensive than building at ground level. For a 100 km long version at 6 g's the muzzle velocity could be as high as 3,460 m/s, but we will use 20 km for this example.

Design Scaling - Assume a 20 km long pipe x 10 m diameter, pushing a 500 ton single stage multiple use rocket. The vehicle will not fill the whole pipe, it is shaped for aerodynamics, and rides on a sled and pusher plate that fits the pipe. It works out the pressure in the barrel needs to be 2 atmospheres (200kPa, 30 psi) to give you 3 g's acceleration, safe for most humans (general public) and satellite parts. Muzzle velocity is 1100 m/s (Mach 3.6), which is not a huge fraction of orbit velocity, but a nice running start before you light up your on-board rocket. Given those starting conditions, a non-cryogenic rocket should have a payload of around 35 tons, which along with a 10 meter maximum diameter should be plenty for any cargo or people you want to launch. This is the upper end of what you might want to build in terms of barrel diameter. For higher mass vehicles, you just need higher operating pressure in the barrel. A first low-g cargo launcher can be a lot smaller than 10 meters, and increased in performance by adding length or going to larger barrels over time. Hawaii is about 20 degrees N latitude, so a launch from there would not be able to reach an equatorial Skyhook, but it would deliver more passengers and cargo than an unaided rocket.

Cayambe, Ecuador

Cayambe is the name of both a city and large mountain about 50 km north-east of Quito, Ecuador. We previously discussed a hypervelocity launcher on the side of the mountain. For transporting people, the barrel will need to be much longer for lower acceleration, and extend west somewhat past the town. For this version we assume a trained crew rather than general public. With pressure suits, conforming seats, +x acceleration (forward facing seats), and crew in good condition you can safely use 6 g's, and thus get a muzzle velocity of 1560 m/s. That's Mach 5.2, or 20% of orbit velocity. The Skyhook has been available since the previous step in the combined system example, which subtracts another 2400 m/s from the rocket stage requirement.



20 km Barrel Fit to Terrain

The **geography of Ecuador** is not a smooth slope like Hawaii. We assume the barrel is 20.25 km long, but curved upward with a segmented radius that keeps centrifugal acceleration at or below 12 m/s^2. That will be felt by passengers as a vertical acceleration (head to toe). The barrel will need to be supported on towers or use tunnels as needed to fit the terrain, and the curvature roughly fits the geography, which is flat initially, rising to a mountain at the end. The ends are at 2778 m elevation south west of the town, and 5731 m at the top of the mountain, with initial and final slopes of 1.4 and 12.4 degrees caused by the curved barrel The gentle curvature keeps the vertical acceleration low relative to the forward acceleration. The higher slope at the muzzle end also allows faster climb through the atmosphere and less drag loss. These assumptions may be changed with more detailed analysis. We assume the rocket stage is 4 x 32 meters in size, and closely fits the barrel, with a mass of 400 tons at launch.

Drag - With a drag coefficient of 0.2, the rocket stage will see 1.93 MN of drag at the muzzle, producing -4.82 m/s^2 deceleration if the rocket does not ignite immediately. The climb rate of sin(12 deg) x 1560 m/s = 335 m/s. The equivalent thickness of the atmosphere is called the **scale height** (8640 m vertically) over which the pressure drops by a factor of e (2.718...). An exponential pressure decay per scale height over many km is how the real atmospheric pressure changes, but it can be approximated as the muzzle pressure for one vertical scale height and then dropping to zero. 8640 m scale height / 335 m/s vertical velocity = 25.76 sec. Multiplied by the deceleration the total drag loss can be estimated at 124 m/s. This value will change depending when the rocket is started, since drag is a function of velocity.

Rocket Performance - The net velocity required for the rocket is found from the Skyhook tip velocity relative to the Earth's center (5074 m/s), less the Earth's rotation at the equator (-465 m/s) and gun velocity (-1560 m/s) plus drag loss (+124 m/s) and other losses and maneuvering which we make an estimate for (+200 m/s). This comes to 3,373 m/s net. The SpaceX Merlin engine has an exhaust velocity of 2980 m/s. 4-6 engines will probably be required for sufficient thrust. The rocket equation gives the rocket mass after reaching the Skyhook as 32.2% of initial mass. Allowing 10% for the vehicle itself gives 22.2% payload, or 89 tons. This is a large passenger and cargo capacity, with a correspondingly large Skyhook to support the arrival mass. A first version would likely be smaller.

Knowing the area of the barrel and the rocket vehicle mass and acceleration, we can calculate the required pressure as 1.91 MPa (277 psi) for the 4 meter gun and 1.22 MPa (177 psi) for a 2.5 meter gun. The challenge will not be barrel pressure, but filling it fast enough when the projectile is moving rapidly. The length will likely require tanks and valves space out along the barrel. The muzzle velocity will likely require a heated gas to fill the pipe, but exactly which gas will be left for detailed analysis. Large gas accelerators have reached above twice the muzzle velocity, so it is more a matter of lowest cost than feasibility.

Spaceport Growth - We had previously built an operational Hypervelocity Gun on the mountain with a muzzle velocity of 5000 m/s and an unaided payload to orbit of 180 kg. With the Skyhook in place, we can calculate the new payload as follows:

- The Skyhook's tip velocity relative to Earth's center is 5074 m/s. Earth's rotation deducts 465 m/s. Drag loss is 1000 m/s from the initial 5000 m/s. Trajectory elevation of 23 degrees means the horizontal component, which is all that counts for getting to orbit, is cos(23 deg) = 0.9205 x 4000 m/s after drag = 3682 m/s. We allow an extra 200 m/s for maneuvering and other unaccounted losses. So the net delta-V of the rocket becomes 1127 m/s.
- Using the same exhaust velocity as the SpaceX Merlin engine (2980 m/s) but at 1/60th the thrust level, we get a final mass of 68.5% x 1200 kg start mass = 822 kg. With the same empty vehicle mass of 180 kg as the version before the Skyhook, we now have 642 kg payload, or about 3.5 times as much.

Going from 642 kg payload with a 60 cm caliber (barrel diameter) gun to 89 tons with the 4 meter caliber human accelerator is a factor of 139 times larger. Since the Skyhook has to be enlarged for the larger delivery mass, a program of gradual improvement will be needed. The launchers will add barrel length and move to larger diameters in steps, and use part of their cargo to deliver Skyhook cable and other materials, so that later deliveries with more payload can be handled. If orbital mining can supply sufficiently strong materials, they can be used, but otherwise they can come from Earth. A smaller version of the human accelerator than the one above could use a 2.5 x 20 m size rocket vehicle with a mass of 100 tons. Using similar calculations, we end up with 20 tons net cargo for it. At some point the low-g accelerator would be too small for seated human passengers, probably around 1.6 meter diameter, but they can still be used for sensitive cargo. Bulk non-sensitive cargo will always have a cost advantage because the higher muzzle velocity lets you deliver 3 times more payload as a percentage of rocket vehicle weight, so it makes sense to keep both types of launchers.

Depending on traffic needs, you may want to keep smaller launchers operating in parallel with their larger replacements. In theory you could launch every time the Skyhook passes over in its orbit, which is every 100 minutes, but barrel cool down or other needs may prevent firing a given gun that often, so having several may be useful. At the upper bound, delivering 89 tons per launch x 14.4 orbits per day x 300 days per year (allowing some maintenance time) yields an astounding 384,480 tons/year to orbit. This compares to the ~1,000 ton/year capacity of current and near-term launchers worldwide.

Cost - At this point, cost has not been estimated to any degree of accuracy. The Falcon 9 rocket has a total mass of 333.4 tons and a payload to low orbit of 10.45 tons. So the ratio of rocket mass besides payload to payload is 30.9 to 1. The bulk cargo launcher has a non payload mass of 558 kg vs 642 kg payload, or a ratio of 0.87 to 1. This 35.5 to 1 advantage should lower costs significantly, but not in that exact ratio. The gun and Skyhook are large installations relative to the rocket stage, and their cost per use will depend on how many times they are used. The Falcon 9 hardware is not currently reused, while the rocket stage is intended to be used multiple times. De-orbiting from the Skyhook is 63% of the unaided velocity from orbit, and thus (0.63)^2 = 39.5% of the kinetic energy to dissipate. This makes the heat shield easier to design, and the stage is pretty rugged in design, since it needs to be fired out of a gun at high acceleration. So in principle it should be able to be recovered and used again.

In the absence of more detailed estimates, for now we will adopt the 35.5 times reduction in rocket size per payload and apply it to the $54 million/10,450 kg = 5,167 $/kg Falcon 9 cost, to get a first estimate of 146 $/kg. To compare to some popular consumer items, the iPad 3 64 GB costs 583 $/kg including packaging, and a Toyota Camry is about 15 $/kg, although neither is designed to survive high-g launch.

Rocket Replacement Alternatives=

Continued on page 2 →

[still to be merged]

Section Header 4

For the self-build options we do preliminary designs, then compare to the existing launcher choices. We need to make some design assumptions to start with:

- **Payload Mass** - We will assume that 20 kg is sufficient mass for a functional hardware item using modern technology. That might need to be changed with a better understanding of payload needs, but we will use it as a starting point. Larger devices can be assembled from several items in orbit, but keeping the item size small lets you use a smaller launch vehicle, and thus lower development cost to start. There is also the possibility to use this as an "express package delivery" service between larger launches on other vehicles, and bring in some revenue.
- **Launch Rate** - We assume an initial rate of about 1 launch per month, and continuing on a steady basis.
- **Production** - We assume one or more **Advanced Manufacturing** type factories, as described on the previous page, are used to build the launcher. This imposes production capabilities on the factory and links the systems. Any materials or components that are not reasonable to make within the factory are bought. The cost of the factory has to be included when deciding which launcher to use.

There are multiple possible ways to launch a small payload to orbit. The conventional approach would be to design a small rocket with two or three stages. Any alternative ideas can be compared to that to see if it has a lower expected development and operating cost.

Low-G Transport

Not every type of cargo can withstand the high acceleration of the Hypervelocity Launcher. In particular, humans are limited to about 3-6 gravities. So until this step of the combined system, we used whatever existing launchers were available for people and delicate cargo. With the availability of the Skyhook we now consider the alternatives for this task and how to choose among them.

Choosing among Alternatives

There are already existing rockets for launching cargo and people into space, and new ones are under development. We assume this will continue to be true in the future. In engineering design, parts are subjected to a **Make or Buy Analysis** to determine whether to make a part internally, or buy it from someone else. This method also applies to the human transport job. If building our own launcher is sufficiently better on cost, technical risk, and other parameters, then we do so. If better alternatives are available from others, then simply buy the launch services. Comparing all the alternatives is called a

Trade Study. First you choose the parameters to use to compare with, and a scoring system to convert diverse parameters into a common scale. You then make the best estimates for each alternative, and choose the one that scores best. When doing a trade study, it is important to use the same assumptions, such as material strengths, for all the alternatives.

Technology and what alternate systems are available changes over time. There is also uncertainty in the engineering estimates before design and construction is completed. Input assumptions like tons of cargo per year can vary over time. Thus for a complex system, a single point comparison is not sufficient. A **Sensitivity Analysis** looks at variations of parameters and assumptions ahead of time to see how it affects the final choice. This can be done efficiently with a mathematical or computer model of the system. Later on, when one of the conditions just mentioned changes, the trade study should be repeated to see if the previous answer is still valid.

Section 4.4: Phase 2B - Industrial Locations for Space (page 2)

← Back to Page 1

Hypervelocity Launcher

The Hypervelocity Launcher is one of the rocket replacement alternatives. It is based on the **Particle Bed Heated Gas Gun** from Section 2.2. In order to replace a large part of conventional rocket propulsion, high accelerations are used. This makes it suitable for bulk cargo, but not for people or complex equipment. Gunpowder-based **Guns** have a long history. **Light-Gas Guns** use lower molecular weight gases instead of the explosive products of gunpowder to reach higher velocities. They have been used for high-speed research since about the mid-20th century. Use of such guns for launch to orbit has been proposed several times in the past, but has not yet been put into practice. Therefore they would require R&D work to go from existing research gun experience to versions capable of orbital launch. We refer to this alternative as a Hypersonic Launcher because the gun muzzle velocity is greater than Mach 5, and therefore the projectile flies through the atmosphere at **Hypersonic Speeds** We choose the particle-bed version because it is relatively simple and has good performance, although there are a number of other gas gun designs.

After establishing an Advanced Manufacturing capability, the next step is a low cost launcher, mainly for bulk materials. Examples of bulk materials are fuel, water, oxygen, structural components, and even frozen food. If there is enough cost advantage in this launcher, you can purposely design parts for the higher acceleration. For example, electronics can survive high g's if they are mounted properly, but most commercial units are designed for lowest assembly cost, not high g's. In general the acceleration of this type of launcher goes down with size. For example, compared to the US M777 field artillery, it may have 5 times higher muzzle velocity, but 150 times longer barrel, thus the acceleration is lower. In the limit of the largest practical gun the acceleration would be low enough for humans (60 m/s^2 or 6 g's), but this initial launcher is at the other end of the size scale, and thus relatively high g-level.

Conventional rockets are used in parallel with this launcher for the balance of delicate cargo and humans which cannot withstand the high acceleration. The particular type of launcher selected is 17 Particle Bed Heated Gas Gun, which is within current technology, has the lowest development cost to start delivering cargo, and a considerable cost advantage over current rockets. If conventional launch costs get low enough, though, this step may be eliminated.

Launcher Scaling

The largest known hypersonic light gas gun was the Lawrence Livermore SHARP Gun in the early 1990's, which reached 3 km/s with a 5 kg projectile, for a kinetic energy of 22.5 MJ. The largest hypervelocity gun launcher was it's namesake predecessor, the High Altitude Research Project (HARP) in the 1960's. That gun was made by welding two 16 inch battleship guns in series, and was able to fire 250 kg at 2300 m/s[1] using a guncotton charge. This had a kinetic energy of 660 MJ, and could have put measurable payload into orbit given suitable propulsion on the projectile. The largest known gun of any kind by energy was the Gustav type German railway siege guns of World War II. It fired a 4800 kg projectile at 820 m/s, for an energy of 1.6 GJ.

For development purposes, it is not good to make too large a jump in scale if you are going beyond past experience. Since the SHARP gun was the largest of that type ever built, we will assume a prototype of about 5 times the energy at 100 MJ (12.5 kg projectile at 4 km/s). This would be followed by a gun large enough to deliver a useful payload to orbit, which we will assume is 10 kg to a 250 km altitude circular orbit. Beyond that, larger guns would be sized by expected cargo traffic. Detailed analysis may change these numbers, but we at least need a starting point to design to. Please note all the following calculations are only preliminary to show how the scaling is performed. For more detailed work, refer to a manual such as Interior Ballistics of Guns, which accounts for more of the real world factors, or do a computer simulation.

Prototype Gun

The location of the prototype gun is not critical. It's main purpose is to reduce the unknowns before designing the orbital gun. For convenience we will assume the White Sands Missile Range as the landing point of the projectiles, since it is large, empty, and designated for missile testing. The gun will then be placed on a mountain slope a suitable distance away, avoiding any population centers under the projectile path. Any other combination of launch and landing points meeting similar conditions will work. We will further assume intermediate physical dimensions and characteristics between the SHARP and Orbital guns, unless there is some technical reason to choose otherwise.

Design Inputs:

- Projectile Mass = 12.5 kg
- Muzzle Velocity = 4000 m/s

Initial Assumptions:

- Barrel Length = ~200 m - This is a geometric mean between the 50 m SHARP barrel, and estimated 800 m orbital gun. May be updated by later calculations.
- Barrel Elevation = 12 degrees - This is based on typical mountain slopes north of White Sands
- Projectile L/D = 8 - This is the length-to-diameter ratio based on a cylinder shape. The actual shape will be to minimize drag, so conical at the front. You want to minimize area to lower drag and barrel size, so a long and skinny projectile, but not so skinny that bending becomes an issue. 8 is a reasonable starting point before structural analysis.
- Projectile Density = 1 g/cm^3 - This is the density of water, a likely filler of test projectiles, and similar to the density of LOX/Kerosine, a likely fuel for an orbital gun projectile.

Derived Values:

- Projectile Diameter = 12.5 cm. - Cylinder volume is pi*r^2*h, and we have assumed that h = 8D = 16r. Thus v = 16*pi*r^3, and we have v=12,500 cm^3 from the mass and density. Solving for r gives 6.29 cm, and we round the diameter of 2r to 12.5 cm. The projectile diameter is also the barrel diameter, if you add a small tolerance for a sliding fit.
- Projectile Acceleration = 40,000 m/s^2 (4,000 g's) - This compares to the peak acceleration of 640,000 m/s^2 of the SHARP gun and 60,000 m/s^2 for conventional artillery. Muzzle velocity is sqrt(2*a*d) where a is the average acceleration and d is the barrel length. Solving for a gives the quoted value. Because gun efficiency falls off at high velocity, we assume the *peak* acceleration is 25% higher than the average acceleration, thus 50,000 m/s^2.
- Peak Pressure = 51 MPa (7,400 psi) - Barrel area is 0.01223 square meters, and gas pressure has to produce a force of F = m * a = 12.5 kg * 50,000 m/s^2 = 625,000 Newtons (N). Pressure is Force/Area. Note this is much lower than the SHARP gun peak pressure of 400 MPa.
- Projectile Range = 54 km - If the barrel slope is 12 degrees, then the projectile is rising at 830 m/s as it leaves the barrel. The density of the Earth's atmosphere decreases with altitude, but the equivalent thickness at constant pressure, known as the **scale height** is about 7.5 km. At a 12 degree elevation, the total path through the atmosphere then is equivalent to 37.5 km. We can find drag on the projectile from the formula F(D)= 0.5* CD * rho * A * v^2. CD is about 0.15 for a conical hypersonic projectile. Rho is air density, which for a starting altitude of 2000 m is about 0.95 kg/m^3. Area and initial velocity of the projectile are given above.

The initial air drag is about 14,000 N, giving a negative acceleration of 1119 m/s^2. This is a significant fraction of the initial velocity per second, so the actual trajectory needs to be found by numerical integration (ie a spreadsheet) using small time intervals so the changes in velocity and drag per time interval, and thus the errors, are small. When this is done the projectile range is found to be 54 km, after a 63 second flight, and reaching a peak altitude of 7350 m above sea level. The impact velocity with no landing devices is about 400 m/s. The small size of the projectile and low gun elevation means air drag has a severe effect on it's path. This is not a problem for the prototype gun, since we are mainly testing the gun. A range of 54 km means the projectile can land within the White Sands Missile Range and not endanger the public.

Orbital Gun

The prototype gun was mainly concerned with demonstrating the components of the gun function properly. The orbital gun demonstrates a larger version, and additionally a functioning projectile that can deliver a small payload to orbit. The gun energy of 1.18 GJ is about 12 times larger than the prototype.

Design Inputs:

- Net Payload to Orbit = 10 kg to 250 km
- Barrel Elevation = 23 degrees - As noted below under Location, Nevada Cayambe is the best location, and we use the actual slope of the west side of the mountain between 4200 and 4600 m elevation. Above that altitude is a

glacier, so we try to stay below that.

- Projectile L/D = 8 - Use the same value as for the prototype gun. An actual projectile design will be needed for for a better estimate.
- Projectile Density = 1 g/cc - Use same value as for the prototype gun.

Initial Assumptions:

- Barrel Length = 800 m - This is an initial guess at a reasonable number. To determine the real length requires a detailed enough design in a form you can vary barrel length, see how the changes affect the rest of the gun system, and then find the optimum value. This approach is called variation of parameters or system optimization, but we need a lot more design detail to attempt it. For now we pick a reasonable starting point.
- Muzzle Velocity = 4500 m/s - This is a reasonable guess based on past gun launcher work. It will also be subject to optimization later.

Derived Values

- Projectile Acceleration = 12,650 m/s^2 (1290 g's) - Found by the same method a for the prototype gun above. Again, this is the average number, so peak acceleration is estimated to be 15,835 m/s^2 (1615 g's). Note this is about three times lower than the prototype gun, mainly because of the longer barrel.
- Peak Pressure = 33.88 MPa (4910 psi) - Found by the same method as the prototype gun. Note the peak pressure is lower by about 1/3 relative to the prototype gun, mainly because of the longer barrel.
- Projectile Mass = 122.5 kg - With a known muzzle velocity and slope, we can try various projectile masses (in the next several paragraphs) to find the one that gives us 10 kg of payload to orbit. The projectile mass can be divided into three main parts: payload, fuel, and empty vehicle. The latter includes all the components like guidance electronics, besides fuel tank and payload support. The on-board rocket is assumed to have an exhaust velocity of 3.3 km/s, typical of a good LOX/Kerosine engine in vacuum.
- Drag Loss = 1725 m/s - As a first approximation, assume that drag losses equal 1200 m/s, the rotation of the Earth at the equator is 465 m/s, and the sea level orbit velocity plus energy of 250 km altitude to be provided by the rocket + gun is 8065 m/s. If the net velocity after drag loss is 3,300 m/s, that leaves 4300 m/s for the rocket. From the rocket equation, the net mass after the rocket burn is 27.2% of the total. We make a first estimate of the vehicle empty hardware mass of 15% based on past rocket designs. Then the net payload is 27.2% - 15% = 12.2%, and the original mass is 10 kg/12.2% = 82.2 kg. Assume the projectile density is 1.0 g/cc. Then it's volume is 82.2 liters (0.0822 m^3), and can be approximated by a cylinder 23.5 cm in diameter and 188 cm long. Since the projectile structure is subjected to a known acceleration, we can estimate the structure mass from the load it sees.

A mass of 82.2 kg subjected to 15,835 m/s^2 peak acceleration requires a force of 1.302 MN. Graphite composite can be assumed to have a strength of 600 MN/m^2, with a density of 1.82 g/cc. For the given load we need 1.302/600 square meters of structure = 21.7 square cm. The forward parts of the structure only have to support what is ahead of that point, so we assume the structure averages 65% of the area over the length of the body. Thus the total structure will be 21.7 x 65% x 188 = 2651 cc, with a mass of 4.8 kg. This is only 5.87% of the total mass of the vehicle, so our 15% assumption for the total empty mass is reasonable.

To get a better estimate of the actual velocity for the rocket, we have to make a better drag loss estimate. Using a spreadsheet trajectory calculator, we find the projectile will fall to 2613 m/s after drag losses have fallen 99%, giving an estimated loss of 1900 m/s. Since this is higher than our original estimate, we recalculate the projectile mass and try again several time until we get a consistent answer. This is called **converging to a solution** After 5 iterations we can estimate the final result is a projectile mass of 122.5 kg, and a drag loss of 1725 m/s.

- Projectile Diameter = 27 cm - this is found from the mass and formulas above. Projectile and barrel areas are both 0.05725 m^2.
- Projectile Range = 615 km - This is the distance the projectile will travel if the rocket does **not** ignite. This is found by following the trajectory calculator until the altitude reaches ground level again. When choosing a launch site, be aware of the impact point of a failed ignition. In this case, the impact point would be near Chiribiquete National Park in Colombia, reached after a 4 minute flight, at a terminal velocity of 1500 m/s if the projectile reaches the ground intact, which is not assured with a full load of fuel and aerodynamic heating. If the rocket engine ignites and then

stops before reaching orbit, the impact point will be on the Equator somewhere east of the ballistic range, with reduced amount of fuel.

Orbital Gun with 2 Stage Projectile

This is an alternate concept to deliver the same payload as the previous design. It uses a 2 stage rocket to see if it improves the overall size. The gun inherently has about 50% higher "exhaust velocity" than the rocket engines, so we divide the total velocity into 3.5 parts to equalize the "difficulty" for each stage. From the previous design, we have a total mission velocity, including drag losses of 9325 m/s, therefore the gun would do (1.5/3.5)* 9325 = about 4000 m/s. Each rocket stage then has to perform about 2660 m/s.

Using our previous assumptions of 15% empty hardware weight and 3.3 km/s exhaust velocity, we can calculate the stage masses as follows:

- Mass Ratios/stage = 2.241 - found from rocket equation. This implies final mass = 1/mass ratio = 44.63% of start mass.
- Stage 2 payload = 10 kg = (44.63% final mass - 15% empty weight) = 29.63% stage 2 start mass.
- Stage 2 start mass = 33.75 kg - From 10 kg / 29.63%.
- Stage 1 final mass = 44.63% start mass, Stage 1 empty weight = (15% x 57.2% fuel used) = 8.31% start mass, therefore stage 2 start mass = 36.32% stage 1 start mass.
- Stage 1 start mass = 93 kg = 33.75 kg / 36.32%. This is significantly smaller than the previous mass, and the gun velocity is assumed to be 500 m/s lower, so we recalculate drag losses using the ballistic calculator and converge to a solution. We end up with 1655 m/s drag loss, and can thus lower the gun velocity by 70 m/s to 3930 m/s.
- Average acceleration is found from $a = v^2/2d = (3930\ m)^2 / 2 * 800m = 9653\ m/s^2$ (984 g's). Peak acceleration is $1.25\ a = 12{,}066\ m/s^2$.
- Projectile diameter is found as above to be 24.5 cm, and barrel area is $0.0473\ m^2$.
- Peak pressure is found from P = (mass x acceleration)/area = $(93\ kg \times 12066\ m/s^2)/0.0473 = 23.72$ MPa.

The combination of lower barrel area and lower pressure leads to a total barrel mass of 57.85% of the previous version. We assume other parts of the gun will scale along with the barrel. The projectile mass is 76% of the previous version. Whether the added complexity of a second stage outweighs the system size reduction is not known.

Operational Gun

The operational gun is designed to deliver paying cargo to orbit. The actual size will be set by how much customer traffic is expected, but for discussion purposes we will assume a 1200 kg projectile launched at 5 km/s. The location is the same mountain as the previous gun. The kinetic energy of this gun, 15 GJ, would exceed the largest previous gun of any kind by nearly 10 times.

Design Inputs:

- Projectile Mass = 1200 kg - Set by assumption.

- Projectile L/D = 8 - As in previous sizes.
- Projectile Density = 1 g/cc - As in previous sizes.
- Muzzle Velocity = 5000 m/s - also set by assumption. This is towards the upper range for light gas guns.
- Barrel Length = 1600 m - This is set by the distance from the bottom of the mountain slope to the glacier line. For this size gun it is likely not worth the extra difficulty building through the ice layer.
- Barrel Elevation = 23 degrees as previous size.

Derived Values:

- Projectile Dimensions = 57.5 cm diameter x 460 cm long - Calculated from mass and density as in previous sizes.
- Projectile Acceleration = 7810 m/s^2 (795 g's) - Again found by same calculation as previous sizes. Allowing for 25% peak increase, this gives 9765 m/s^2 peak acceleration (just under 1000 g's). We want lower acceleration for larger projectiles both to keep the barrel pressure reasonable, and less load on the projectile structure with the larger mass.
- Peak Pressure = 45.1 MPa (6545 psi) - This is slightly higher than the previous size. Note that in a large gun with a pressure drop as it fires, only the bottom end will see the peak pressure. The muzzle end can use a lower strength pipe. Using high strength steel for the barrel, the peak pressure requires a barrel wall thickness of around 6 cm, which is reasonable.
- Empty Vehicle = 180 kg - The projectile is sufficiently larger that we should check the empty weight rather than assume the previous percent fraction. The peak acceleration force is 11.7 MN, giving a structural area of 195 square cm at the base, and a total structure volume of 58,300 cc. This comes to 106 kg mass, so our assumption of 15% for total empty vehicle (180 kg) is still reasonable.
- Drag Loss = 1000 m/s - This is found by using the trajectory calculator to the point you are above 99% of the atmosphere (34.5 km above sea level)
- Payload = 180 kg - The velocity to reach orbit is 8065 m/s. Subtracting the rotation of the Earth (463 m/s) and the remaining velocity after drag loss (4000 m/s) gives 3602 m/s to be added by the projectile. The rocket equation gives a remaining weight of 360 kg. Subtracting the empty vehicle leaves 180 kg (400 lb) of net payload.

Note that the projectile mass increased by about 10 times, while the payload increased 18 times over the previous size. This is due to higher muzzle velocity from the gun, and lower relative drag losses from the larger projectile.

Large Gun

This gun is towards the upper end of length that can be built on Cayembe's slope. The diameter is set to 1.2 meters. This assumes there is enough bulk cargo traffic to justify the larger gun size.

Design Inputs:

- Projectile L/D = 8 - As in previous sizes.
- Projectile Density = 1 g/cc - As in previous sizes.
- Muzzle Velocity = 5000 m/s - As in previous size.
- Barrel Length = 3200 m - This is the upper end of length that can be fit to the mountain slope. The bottom end is at 3960 m elevation, and the upper end is at 5300 m, giving a 1340 m rise. The mountain slope is not as constant at the ends because we are maximizing length, so some of the barrel will need to be supported above ground level. Also this extends above the glacier line, so the barrel will either have to be supported above the ice, or protected from ice movements.
- Barrel Elevation = 24.75 degrees - Found from simple trigonometry of arcsin(rise/barrel length).

Derived Values:

- Projectile Dimensions = 1.2 m diameter x 9.6 m long - From design inputs above.
- Projectile Mass = 10,850 kg - from dimensions and density above.
- Projectile Acceleration = 3905 m/s^2 (400 g's) average - Again found by same calculation as previous sizes. Allowing for 25% peak increase, this gives 4880 m/s^2 peak acceleration (just under 500 g's). By keeping the muzzle velocity the same, but doubling the barrel length, the accelerations are halved.
- Peak Pressure = 46.8 MPa (6790 psi) - This is about the same as the previous size.
- Empty Vehicle = 1625 kg - We again check the empty weight rather than assume the previous percent fraction. The peak acceleration force is 52.95 MN, giving a structural area of 882.5 square cm at the base, and a total structure volume of 423.5 liters. This comes to 771 kg mass, so our assumption of 15% for total empty vehicle (1625 kg) is still reasonable.
- Drag Loss = 460 m/s total velocity - This is found by using the trajectory calculator to the point you are above 99% of the atmosphere (34.5 km above sea level). Horizontal velocity component is 4,160 m/s at the peak of the ballistic arc, assuming the rocket stage does not fire, and altitude will be 194 km at this point.
- Payload = 180 kg - For this size gun we assume a Skyhook from a later step is available. Skyhook tip velocity relative to the Earth's center is 5074 m/s. Subtracting the rotation of the Earth (465 m/s) and the remaining projectile horizontal velocity after drag loss (4160 m/s) gives 449 m/s to be added by the projectile. The rocket equation gives a remaining weight of 9330 kg. Subtracting the empty vehicle leaves 7,705 kg (17,000 lb) of net payload.

Note that the projectile mass increased by about 10 times, while the payload increased 43 times over the previous size. The additional 4 times gain beyond projectile size is due to the Skyhook. With a slight improvement in gun velocity or Skyhook velocity or orbit altitude, the projectile would not have to add any velocity, only do maneuvering to meet the Skyhook landing platform. In that case the net payload would rise another 20% to 9,225 kg.

System Design

With the scaling of the launcher and projectile completed, you can now do a preliminary design. Preliminary design stops just before you start the final drawings and calculations, and is in enough detail that you can do cost estimates, and tell if any parts require new research.

Location

As noted above, the location for a sub-orbital prototype is not critical. For a full orbital gun, the best location on Earth is probably Nevada Cayambe, the 3rd highest mountain in Equador, and highest point in the world on the Equator at 0 deg N, 78 deg W, 5790 m above sea level. It has a nice slope (23 degrees elevation) pointing east just below the snow line, and you can position a gun so the terrain slope past the muzzle tends to decrease. This is obviously critical so projectiles don't hit the mountain itself. For now there is a glacier on top of this mountain, even though it sits on the Equator, so the uppermost 1200m of the mountain would be difficult to build on. Launching from 4600 m saves you from going through 46% the atmosphere, which significantly cuts drag losses and heating, and an equatorial space station as a destination will pass overhead every 90 minutes or so. Any launch site off the equator will be limited to one or two times a day to launch. Other places will work as a launch site, just somewhat less efficiently.

Because of it's rotation, the Earth is slightly fatter at the equator. So even though there are taller mountains measured from sea level, equatorial ones are "higher" in terms of reaching orbit, and get more benefit from the Earth's rotation. The latter can be found from the Equatorial radius (6378.1 km) plus launch altitude (4.6 km), divided by the time it takes to rotate once, which is called a sidereal day. This is 86,164 seconds, thus the rotation amounts to 465 m/s at this location, or 5.77% of the total velocity to reach orbit.

Other considerations for a launch site are distance to populated areas, since the gun is very loud, avoidance of avalanche zones, and optimum slope. To reach orbit you want the maximum kinetic plus altitude energy remaining after drag effects. Some barrel elevation will be optimum for this, and finding a mountain with a matching slope will minimize construction cost. The alternatives of tunneling or building a support structure for the barrel would be much more expensive. To find the optimum elevation for a given gun design, use the trajectory calculator and input different angles to find the best one.

Barrel

The barrel for this type of launcher is basically a large high pressure pipe. The most similar industrial item is a natural gas pipeline. For cost reasons, the majority of the barrel strength will likely be high strength steel. Since the inside will be exposed to hot hydrogen gas for a short period, and some amount of wear from projectile friction, it may need a liner or coating. The need for that will be determined by **materials and thermal analyses**, which are a standard part of mechanical engineering. The gas and projectile do not produce great force on the barrel lengthwise, the reaction force to the gun firing will mostly be at the back end, which is called the **breech** in gun terminology. Even so, the barrel will likely be made of bolted sections, which allows removal for maintenance, mounting and alignment to the ground, and sliding fit sections for expansion from general weather changes and heating from firing.

The upper end of the barrel will likely have a flap system to keep most of the air out of the barrel. The flaps are pushed open by the remaining air piled up ahead of the projectile. A "silencer" type device may be needed at the muzzle, either to actually lower the acoustic levels, reduce the muzzle flash from the hot hydrogen burning when it meets air, or to capture the Hydrogen gas to be used again. Finally, a shaped nozzle or steam ejector at the muzzle may be needed to ease the change from high acceleration in the barrel to negative acceleration from air drag once outside it. These options should be analyzed and tested with the prototype gun.

Heat Exchanger

The heat exchanger stores heat from a convenient source (probably gas burners or electric heating elements), then transfers it quickly to the Hydrogen gas when the gun fires. Aluminum Oxide particles, commonly used for sandpaper, are used as the storage medium. They can withstand high temperature, and in grain form can transfer heat quickly due to the large surface area.

Breech

This part of the gun is where projectiles are loaded, fueled if needed, and the main reaction force of the gun is passed into the ground.

Storage Tanks

The Hydrogen needed to fire the gun is stored at room temperature in ordinary high pressure tanks. The storage pressure in the tanks is higher than the operating pressure in the barrel so that the gas will flow correctly, but the tank volume will be smaller than the barrel because room temperature gas takes less volume than the same amount of hot gas.

Low Density Tunnel (Optional)

- See also: Low Density Tunnel

In theory you could eliminate the drag and heating of climbing through the atmosphere by using a vacuum tunnel beyond the end of the barrel. This would be larger diameter than the barrel, to eliminate friction, and continue upward as high as drag savings and cost dictate. In practice, you can get 93% of the effect of a vacuum tunnel with a hydrogen tunnel. Drag is proportional to gas density, and hydrogen gas is 93% less dense than air. It also tends to float, so that simplifies holding the tunnel in the air above the end of your launch mountain. Also, hydrogen is your gun propellant, so it just mixes with the hydrogen in the tunnel, and it becomes fairly easy to pump back for the next launch.

You build a large enough tunnel so the hypersonic shock waves of the projectile won't destroy it. That also gives more lifting volume. How far you build the tunnel, or whether to use it at all comes down to cost. It can be added later to a basic gun to hold down initial cost. With such a tunnel in place, you can fire at higher velocities as well as lower the overall losses.

Projectile Design

This example of the projectile design is for the operational gun with a 1200 kg projectile. The other sizes will be smaller and simpler versions will lower loads, so this is a "worst case" design challenge to solve.

Structure

Propulsion

Thermal and Re-entry

The projectile needs to survive three periods of high heating: (1) Within the barrel from hot Hydrogen gas pushing it, (2) While flying up through the atmosphere at high speed, and (3) during re-entry so you can recover and use it again.

The first heating source comes from behind. Hypersonic guns often use a **Sabot**, a disposable structure to help fit the projectile to the barrel and provide a better seal for the gas pressure. Using a sabot in this design can protect the projectile from heating either by including insulation, or thermal inertia, which is simply that solid objects take time to heat up. If that time is longer than it takes to leave the barrel, then extra insulation is not needed. The alternative is to use the rocket nozzle at the back of the projectile, which already needs to withstand high heating. In that case it also needs to withstand the high acceleration force from the gas pressure.

The second and third heating sources come from the front. Flying up through the atmosphere at high velocity causes the higher peak heating rate, but re-entering from the even higher orbital velocity gives a higher total heating, although at a lower rate. The common method to deal with high heating rates is an **ablative heat shield**, which decomposes and generates a protective gas layer. To keep the projectile pointed in the right direction both flying up and during re-entry may require fins or **control surfaces**, which would require heat shielding on them.

Thermal protection is measured in terms of the amount of energy it has to dissipate. While flying up through the atmosphere, the projectile goes from 5 to 4 km/s via drag, thus losing 5.4 GJ of kinetic energy. During re-entry, when the projectile is empty, it dissipates 5.2 GJ. In the former, the heat pulse starts at the highest rate, decaying exponentially as the air pressure goes down with altitude with a time constant of about 4 seconds. In the latter, the heat pulse starts slowly as the projectile encounters the thin upper atmosphere at high velocity, reaches a peak at lower, thicker air density but still fairly high velocity, and then tapers off as velocity falls faster than pressure rises. The exact duration will have to be found by trajectory simulation, but is measued in minutes rather than seconds.

The projectile structure is designed to withstand 800 g's at launch, and therefore in theory should be able to handle a similar deceleration at landing. In reality, empty fuel tanks can buckle more easily than full ones (think of a full vs empty soda bottle for example), but the projectile is still a sturdy device. **Terminal velocity** is the velocity a falling object will reach where drag equals gravity, and so velocity stops changing. For our empty projectile, we can calculate terminal velocity assuming rear control surfaces add 50% to the total area and have a drag coefficient of 0.4. Since the projectile has an empty mass of 180 kg, gravity produces 1764 N downward force. The terminal velocity is then about 150 m/s. Assuming the empty vehicle can withstand 2000 m/s^2 (200 g's), it can reasonably sustain an impact velocity of around 20 m/s using crushable structures. Therefore it needs a device like a parachute or much larger control surfaces to create drag. Forcing the projectile to land sideways instead of nose-first will increase drag area and drag coefficient and lower teminal velocity.

If the launch site is on the Equator, then the landing point will also be on the Equator so long as the projectile propulsion is always aligned east-west. In this example, the landing point can be the relatively unpopulated areas about 80 km east of the Andean mountain chain, which makes return of the empty projectile to the launch site relatively easy. Then it is a matter of timing the re-entry burn of the on-board rocket to set the landing point. The size and weight of the empty projectile can be handled by a small truck.

Navigation

Cargo Deployment

We assume the cargo being delivered has the same average density as the rest of the vehicle, 1 g/cc. Allowing for the projectile structure, the cargo diameter is set to 50 cm. Given a mass of 180 kg, then the length is 92 cm. Depending how the structure needs to be designed, the cargo opening can either be a side hatch, or the end of the projectile can hinge open and the cargo exits that way. Two possibilities for delivering the cargo are (1) without any station or depot, and (2) with the assistance of a station or depot. On the first case, the projectile only needs to maneuver to the accuracy of the desired payload orbit. In the second case, it needs to maneuver close enough for the station or depot to rendezvous. The final docking can be done either by the projectile or station. It likely will be more efficient to put the rendezvous system once on the station, rather than launching it each time with each projectile. Given the size of the projectile, simply grasping itwith a clamp around the middle may be the easiest way to dock.

If the cargo is liquid, an integral (built in) tank will probably save weight and deliver more mass. Then you will have two projectile designs, for liquid and dry cargo. The projectile can be designed with replaceable cargo modules for liquid, dense, or normal cargo. The weight savings for custom modules will have to be compared to the design and operationscost penalty for having multiple designs.

Traffic and Schedules

Performance, Cost, and Risk Estimates

Section 4.5: Phase 4A - Low Orbit Development

Low Earth Orbit (LEO) was the first region in space to be developed, starting in 1957 with the launch of the first orbital satellite. It still accounts for the largest segment of artificial mass in orbit, but it is not yet a complete economy. As of 2017 the activities there are mainly public programs, such as scientific and military, and support of activity on the ground, such as Earth observation and some communications. The limited range of activity is a function of the high cost of reaching orbit compared to most activity on the ground. Industrial development on Earth in Phase 2B should greatly reduce transport cost, and exploiting materials and energy already in space should further reduce total costs. The combination should enable larger scale development and a wider economic base in the region.

For the purpose of our program we consider Phase 4A to have started once industrial development on Earth (Phase 2B) enabled equipment delivery to orbit. We recognize the existing activities, but consider additions or changes that can be made in the future. New activities in this phase will generally precede other phases in space because Low Orbit is the physically nearest and lowest energy region to reach from Earth. However, most of the useful resources in space are beyond Low Orbit. So a large part of the development in this region will be to support the more distant regions. In turn, the Low Orbit region needs continuing support from Phase 2B industries for space, because industrial-scale production and transport systems are needed to reach low orbit. This phase, like others, will continue in parallel once started.

The Low Orbit region is already used for a number of purposes, and quite a few additional ones are possible in the future. This makes it a complex engineering challenge. Therefore we apply the Systems Engineering methods from **Section 1.5**. That begins with concept exploration, which we pursue in this section. We first describe the features of the region and do an industry survey to identify additional future activities. We then look at drivers like motivations, economics, and technology, which would cause projects to be initiated. This information is combined into a development approach and a set of activities and projects by time and function. Finally we link the projects to each other and other program phases. Projects for which we have developed additional details and calculations are then more fully described as the last part of the section. As an output from our concept exploration, we identify what R&D will be needed to prepare for this phase, and feed it back to the preceding Phase 0G - R&D for Low Orbit Development.

Low Orbit Features

Region Definition

Our program defines the **Low Earth Orbit** (LEO) region as orbits averaging 160 to 2700 km above the mean radius of the Earth. When it would not be confused with low orbits around other bodies, we will often shorten this "Low Orbit". It is entirely surrounded by the High Earth Orbit region, and in turn completely surrounds the Earth. It would be very difficult to build static structures extending up from the surface to more than 160 km, and motionless objects would rapidly fall if unsupported. Therefore physical objects which persist in this region need to be in motion with a particular range of velocity and direction, such that they do not intersect the Earth or rise too high. This is unlike locations on Earth, where objects can have fixed coordinates of latitude, longitude, and altitude.

Locations in Phases 4A to 4F are identified by a set of **Orbital Elements**. These parameters determine the size and shape of the orbit, how it is oriented in three dimensions, and a position along the orbit at a given time. For Low Earth orbits, the parameters are usually referenced to the center of the Earth and a fixed direction on the **Celestial Sphere**. An object's position constantly changes due to orbital motion, but can be projected from the given into the past and future. Orbit parameters can change over time, either from natural forces or using any of the transport methods listed in **Part 2**. So a built-up site, like a space station, may change its orbit over time.

The lower bound of the region is set by where atmospheric drag would rapidly cause orbital decay without compensating propulsion. This is somewhat higher than the 80-122 km **Boundary Designations** between the atmosphere and space determined by other methods. The 80-160 km altitude range has space-like conditions, such as near-vacuum pressure levels, but it can only be transiently occupied by unsupported objects. More permanent occupation requires attachment to either the ground or objects in higher orbits. We therefore assign items in this transition range to earlier phases if transient or attached to the ground, and to this phase if attached to objects above 160 km.

The region's upper bound is set halfway in energy terms between the lowest stable orbits and Earth escape, or 75% of the energy from the Earth's surface to escape. This is an arbitrary limit, but lower orbits have different enough conditions from higher ones to warrant a distinction. Orbits can be elliptical, where the altitude varies as you move along them, so we define the region by the average of perigee and apogee along the major **Orbit Axis**. The highest point of an elliptical orbit in the region is then 5240 km (0.82 $R_{\oplus}$) above the surface. Since orbits have many possible orientations and shapes, the region has a fuzzy boundary in physical space. Instead, objects meet or do not meet our definition for the region.

The Earth's gravity is the dominant force in the region, at least 74,200 times stronger than than the Moon's and 500 times stronger than the Sun. Total volume of the region using the average upper altitude bound is 1.8 times the volume of the Earth, and cross section is 85.7% of the Earth's land area. Not all of this can be used because of intersecting orbits and blocking sunlight from reaching the surface. Despite this, the usable space is substantial.

Environment Parameters

Designs for long-lasting objects in Low Orbit must accommodate the local environment conditions. We consider the same environment parameters as for projects on Earth and other space regions. Where local conditions are outside the ranges of previous designs, they must be modified accordingly. We also note unique conditions for Low Orbit which must be considered in designs.

Primary Parameters

Temperature

A **Black Body** at the Earth's average distance from the Sun has an equilibrium temperature of 393.7 K (120.5 C) on the Sun-facing side. The movement of objects in Low Orbit are centered on the Earth, and therefore share the same theoretical equilibrium temperature. Actual hardware temperature will be a function of its time in the Earth's shadow, orientation, infrared contribution from the Earth as a function of altitude, albedo, emissivities, and thermal properties. As an example, 50% reflective grey body that has a back side facing away from the Sun would have a temperature of 331.0 K (57.9 C) if it were in sunlight all the time. However, low orbits are typically in the Earth's shadow 22-40% of the time, reducing the temperature to 245 to 262 K (-28 to -11 C) The result for various hardware designs is an average temperature within the moderate to extreme ranges encountered on Earth, but with short-term fluctuations from crossing the Earth's shadow.

Space equipment is currently designed for low mass, and therefore low thermal mass. The vacuum of space also lacks convective and conductive heat transfer with the surroundings. So the internal temperature can vary significantly as the equipment moves into and out of sunlight. Parts of a satellite may also be normally oriented to face towards or away from the Sun, and therefore be hotter or colder than average. So while average temperatures may be reasonable, specific equipment elements may need adapted designs to account for their operating conditions.

Atmosphere and Water Supply

The **Earth's Atmosphere** extends from the surface to about 10,000 km, and therefore fills the entire Low Orbit region. However, the density decreases with altitude, and at 160 km is one billion times lower than at sea level, placing it in the extreme range as far as people, and hard vacuum condition for many design purposes. Dynamic pressure at 160 km due to orbital velocity is less than 0.05 N/m^2. This is 1.8 million times lower than static pressure at sea level and about 180,000 times lower than passenger aircraft at cruising altitude. So it is negligible for most design purposes, although enough to cause orbital decay. Natural water essentially doesn't exist in this region, requiring import from elsewhere. Since transportation costs are currently high, equipment and processes that currently use water would likely need modification to reduce their use, or substitution by methods that don't require it.

Ground Strength

Since this is an orbital region, soil or rock strength for construction or transportation purposes is not relevant.

Gravity Level

The Earth's gravity in the region varies with altitude from 9.3 to 2.95 m/s^2, or 95 to 30% of surface gravity. The variation from 160 to 5240 km altitude goes as the inverse square of the distance from the Earth's center. However, most objects in the region will be in orbits with free fall conditions. Gravity still accelerates them downwards, but their horizontal motion is enough that the Earth curves away a compensating amount. Different parts of smaller objects see about the same acceleration, so they don't see net acceleration among themselves. The effect at small scales is as if there were no gravity acting, and thus no structural loads for design purposes.

The situation changes for larger objects. Their parts are different distances from the Earth's center, and the direction to the center varies. So the gravity forces are different strength and direction, resulting in net forces between the parts. The larger the object, the more noticeable these differences become, to where they can be the primary design load. Artificial gravity is desirable for human health, and for some production and transport methods. This can be produced by rotation, and creates additional design loads.

Radiation Level

The Earth has natural **Radiation Belts** containing high energy charged particles. They mostly come from the **Solar Wind**, and are trapped by the planet's **Magnetic Field**. The inner belt generally extends from 1000 to 6000 km altitude, but parts may reach as low as 200 km. It therefore fills most of the Low Orbit region. The magnetic field, and thus the belts, are generally toroidal (doughnut) shaped, so the radiation levels vary strongly by altitude and latitude. The magnetic field is tilted and off center with respect to the Earth's polar axis, and solar wind pressure causes variable distortions to it. So the belts can move, and radiation levels will vary by orbit position and time.

An unshielded person in the dense parts of the belts can get a lethal radiation dose in a matter of days or months, and the radiation can cause permanent damage and transient upsets to equipment. To date, the main ways to avoid human exposure have been to stay in lower orbits, where their strength is less, and, in the case of Lunar missions, to transit them quickly at higher latitudes. A number of satellites have orbits that require them to be in higher radiation regions of the belts. This requires **Radiation Hardening** in their designs. Future approaches to the radiation problem include local shielding by bulk mass, weakening the belts by intercepting the particles with mass or electrostatic devices, and reducing their particle sources with devices "upwind" between the Earth and Sun.

Communication Time

Most long distance communication in the region would be by electromagnetic waves (radio or laser) in a vacuum. So round-trip (ping) communication time with Earth and within the region is mainly a function of distance and the speed of light. Direct communication with Earth can be as little as 1 millisecond (ms), though that would be rare. From the highest orbit altitude in the region to the Earth's horizon is 9700 km, and therefore 65 ms ping time. There would be additional transmission time from the space-to-ground terminal to the end point of communication on the ground. Direct communication between points in the region can take up to twice that with the ground, for paths which graze the Earth from maximum altitude on both ends, so 130 ms. Direct signals can't pass through the Earth, so some space-to-ground and space-to-space communications will have to go through one or more relay points. Today this is commonly via satellites in synchronous orbit, because it provides a fixed target for ground stations and three satellites can provide coverage for most of the planet. A worst-case link might require two satellites in synchronous orbit at maximum distance, and therefore 1100 ms ping time (1.1 seconds). Low orbit point-to-point satellite networks are under development. They require many more satellites, because each one only sees a small part of the Earth at any time. Such networks would reduce the worst-case ping time to about 180 ms.

Travel Time

Orbit periods in the region range from about 90 to 144 minutes. The Earth rotates beneath given orbits, and objects are in constant motion at different speeds in different orbits. Therefore travel time from Earth to points in the region, or between points in the region, are usually governed by waiting times for proper alignment rather than the orbit periods themselves. The combination of the Earth's rotation and safe launch directions from a given site typically result in one launch window per day. Once reaching Earth orbit, careful matching with the destination may take another day. So total travel time will be 1-2 days, although planning for space travel today involves much longer lead times for training and securing a ride.

The planes of inclined orbits shift due to the Earth's equatorial bulge on the scale of several degrees per day. So minimum energy orbit-to-orbit transits, which require the orbit planes to align, can require on the order of 100 days waiting time. Point-to-point travel between orbits can be accomplished much faster, but at great expense in propellant if using chemical rockets. Electric or other propulsion can be much more efficient, but are themselves typically slow in providing the needed velocity changes. Future traffic within the region and points beyond will have an incentive to concentrate on equatorial orbits. These reduce waiting time for launch, because the orbit always passes over the same points on the ground every time. They also eliminate the plane shift effect from the equatorial bulge. There will still be a need for orbits with other inclinations, so not all traffic will be equatorial. They will still have to deal with travel delays.

Stay Time

Recent missions to the International Space Station, which is in the Low Orbit region, have averaged 6 months, with a maximum of 1 year. By comparison the US average stay time in a given county, which for this purpose is a single location, is 25 years, with the most rapidly growing ones averaging 7 years from growth plus mobility. Therefore orbital stay times are short compared to those on the ground. The short orbital stay times impose extra transportation requirements, but relieve the personal space and comfort needs because the crew understand the conditions to be temporary. Current stay times are limited by radiation exposure and the long-term effects of zero gravity, despite attempts to counteract them. If longer-term stays are desired, then designs would have to address the radiation and gravity problems, and provide increased personal space and comfort. The current ISS is designed for zero gravity research, so those kinds of design changes would imply a new orbital installation.

Transport Energy - The minimum theoretical transport energy from the Earth's surface to the lowest altitude in the region is 32 MJ/kg, accounting for kinetic and potential energy, less the contribution of the Earth's rotation. The highest orbits in the region require about 50% more, or 48 MJ/kg. However, current chemical rockets such as the **Falcon 9** deliver about 4% of takeoff mass as payload, and most of the rest is propellant whose energy is consumed. Assuming 90% of the takeoff mass is RP-1/Oxygen with a chemical energy of 13 MJ/kg, the transport energy consumed is 285 MJ/kg of payload, or a system efficiency of 11%.

Unique Conditions

Day Length

Orbits in this region range from 87.5 to 143.5 minutes, with a day/night cycle if a satellite crosses into the Earth's shadow. Some orbits are "sun synchronous", with their path oriented to avoid darkness, but most are not.

Available Resources

Energy Resources

Solar has been the primary energy source for satellites in the region since 1962. A notable exception was the Space Shuttle, which used fuel cells, and a small percentage of satellites which use **Nuclear Power**. Many satellites have batteries to cover time in the Earth's shadow and occasional eclipses.

Solar Energy - The **Solar Constant** at 1 AU is 1361 W/m^2 at solar minimum and about 1 W higher at solar maximum. The Earth's orbit varies from 0.987 to 1.017 AU from the Sun, which changes the local intensity from +2.65% to -3.3% of the reference value. Orbit radius has a negligible effect on solar distance and flux. Depending on orbit parameters, satellites can spend up to 40% of their time in shadow, generally decreasing with altitude, and varying cyclically due to orbital precession. Total available energy in the region is large (352,000 TW), but it cannot all be used, because it would block sunlight from reaching the Earth. A reasonable limit is 1% of this, which is still 175 times total 2017 energy use by civilization. Sunlight in natural form is useful for illumination and plant growth, and conversion to electricity using solar panels is reliable and up to 30% efficient. Concentrating reflectors can produce high temperatures, and somewhat higher electrical efficiency at the cost of complexity and mass.

Other Energy Sources - Batteries are the usual way to handle the 35 minutes or less of darkness per orbit, and fuel cells and nuclear power have been used in the past. Some future possibilities are beamed energy from the ground or other satellites, and tapping the satellite's orbital energy using electrical conductors or momentum exchange. In order for the orbit not to decay, the orbital energy must be replaced from other sources.

Material Resources

Earth's Atmosphere - The atmosphere above 160 km contains about 32.5 kg/km^2, or 17.5 million tons in total. If some of this gas is collected, more will replace it from below, so it is a large, although low density, resource. The scale height, over which pressure drops by a factor of *e*, at this elevation is 26.4 km, increasing to 36 km at 200 km altitude, so most of gas is concentrated in the 160-200 km altitude range. The composition in this range is mostly nitrogen (N_2), and monatomic oxygen (O) rather than diatomic oxygen (O_2)found at low altitude. Collection of air from this region can be accomplished from orbit with a reverse nozzle that compresses the incoming flow. However, the drag this creates must be balanced with thrust produced with electric propulsion, using a portion of the collected gas.

Orbital Debris - As of 2017, the region contains several thousand tons of non-functional satellites, empty upper stages, and smaller fragments generated by collision and other reasons. The debris mass continues to increase faster than orbital decay removes them. All the artificial debris was designed for use in space, and some of it likely still contains functional parts, even though the satellite as a whole no longer works. It is a potential source of usable materials and parts which are already in the region, although how useful is yet to be proven. At the least, efforts should be made to remove them because they are a hazard.

Other Sources - Other natural sources of materials are negligible, such as fluxes of meteors, dust, and particles, and the Earth has no natural satellites in the region. So other needed materials will have to be imported, either from Earth or more distant regions. Delivery from higher orbits to the region can mostly be accomplished by slow **Aerobraking** against the Earth's atmosphere. This uses the thinner atmosphere at higher altitudes than **Atmospheric Entry**, so it does not produce high accelerations or extreme heating. The remainder of the maneuvers are accomplished by regular propulsion methods, but overall it is much more efficient. Even though the Earth's surface is physically close, delivery from there requires much more energy.

Industry Survey

- Existing satellites, projects & programs already using the regionThey can supply inputs or accept outputs from the phase, or adapted for phase needs
- Proximity to Earth, so products (materials, parts, equipment) & energy (beamed, chemical, nuclear) can be delivered from existing civilization.Labor can be supplied directly or remotely

Project Drivers

Development Projects

General Approach

Current and Near-Term Projects

Long-Term Projects

Low Orbit Production

Low Orbit Habitation

Low Orbit Transport

Orbital Tugs

Crew Transport

Transport Infrastructure

(skyhooks, spaceports, storage)

Low Orbit Services

Program Integration

Concept Details

Continued on page 2 →

[still to be merged]

Concept Exploration Approach

Systems evolve through a sequence of life cycle stages, from initial idea to final disposal. For projects and locations in the Low Orbit region, we will address their early design stages, which are concept exploration and conceptual design, with the following set of tasks, which are further detailed below:

- 1. Concept Exploration

    1.1 Region Definition - Boundaries, environment, and resources
    1.2 Phase Candidates - Activities from program goals & objectives, reference architecture, industry lists, and plans for future space projects

1.3 Phase Needs - Compare to current space programs to identify new projects & locations
1.4 Phase Concept - Organize into a logical sequence & link to other phases

- 2. Develop Reference Architecture
- 3. Identify Requirements & Measures
- 4. Perform Functional Analyses
- 5. Allocate Requirements
- 6. Model Alternatives & Systems

6.1 Collect External Technical Information
6.2 Develop Alternative Options

6.2.1 Identify Relevant Fields by Function
6.2.2 Develop Candidate Technology & Methods List
6.2.3 Assess Candidate Feasibility
6.2.4 Size Relevant Options
6.2.5 Quantify Option Parameters & Configurations

6.3 Build System Models

- 7. Optimize & Trade-Off Alternatives
- 8. Synthesize & Document Conceptual Design

8.1 Write Conceptual Books & Articles
8.2 Write Design Technical Reports

Heading

Heading

Section 4.5: Phase 4A - Low Orbit Development (page 2)

← Back to Page 1

Assembly Platforms

Given a conventional method to get cargo to orbit, and possibly a **Hypervelocity Launcher** (previous section), the next step is an **Assembly Station** to collect cargo from the high gee launcher, and more sensitive cargo and humans from conventional rockets, and build larger systems there for further space projects. Assembly lowers the required size of individual launches, and thus the up-front development cost.

History

Prior manned space stations, especially the current International Space Station, have used orbital assembly as an engineering method. Particular features for assembly include, first of all, designing the parts to be assembled. Mechanical docking devices include guides to align the parts, latches and powered bolts to fasten them firmly, and electrical and other connectors which automatically join when the parts are brought together. The heavy duty physical tasks of assembly are carried out by a rail-mounted robotic manipulator arm, normally controlled by an on-board operator. Lighter duty tasks are done by humans in pressurized suits. The experience from that project is a good starting point, but it has been about 20 years since the Space Station design was set. Computers and communications have advanced a great deal since then. Additionally, design for continuous growth requires a different design philosophy

Design Approach

The approach in this example uses smaller modular components than in the past. In the early construction stages these are assembled with remote controlled/automated robotic arms into larger units. Once sufficient facilities are in place, human crew can be added. The robotic work uses experience from the **Advanced Manufacturing** step. The Assembly Station starts very basic, and gradually extends it's capability by adding more modular parts, and later by manufacturing items locally, rather than just assembling delivered cargo. An important part of the design is to use standardized modular components. That way new parts can be added in any arrangement and they will still fit, and new designs are not needed for each new job. Additionally, use of standard parts makes it easier to stock spares.

Parts like truss elements, which are naturally strong, can be packed for gun launch, then assembled into a complete truss. Pressurized modules with rigid walls would not fit into a gun-launch cargo, but inflatable modules might. Alternately conical or dome sections can be nested and then assembled into complete modules. If a vacuum welder or laminated tape winder were available, module assembly from smaller pieces would be possible. So each component needs to be looked at to find the best method of delivering it, and it will likely end up a mix of launch methods.

The output of the Assembly Station would be commercial items like spacecraft or sections of spacecraft, and also internal production that would extend the range of later steps. For example, the Assembly Station could assemble a mining tug from parts, which then goes to collect materials from a Near Earth Asteroid. The Station could "reproduce" itself in a sense, by splitting off or assembling a subset which can then go off and be the seed for construction in a new location. Initial growth is by simply adding more modules of a given size. Later growth can be by using larger launch systems from Earth when the economics justify it, or by producing larger components for later "generations" of construction.

Component Types

The following is a list of parts for start up of an Assembly Station. It will require a lot more detailed analysis and design to reach a final list, but this will illustrate the types of modules that would go into such a design. First launch may be by another launch system, to enable a complete functioning system to get delivered as a unit. Later launches can be smaller elements as additions delivered part by part.

Thuster Unit

One of the first items to deliver to orbit will be a small chemical propulsion unit. It will include tanks, fuel and small thrusters, and a way to dock firmly to other structures. The docking port may be as simple as a magnet to attract another payload, and then some bolt or clamp to secure it. The propulsion unit does all the moving around to line up with the payload. Docking other payloads will automatically connect power and data lines. For a first launch, it may be feasible to launch an electronics unit and a partly fueled thruster unit as a single cargo. Otherwise a larger capacity launch system is used.

Fuel Tanks

The growing assembly station will use fuel to meet each cargo as it reaches orbit, and also to make up for drag losses from the thin atmosphere that exists at any low orbit altitude. Therefore it will need periodic fueling.

Electronics Unit

This will contain some smaller solar arrays for power, some computer systems, batteries, one or more cameras and GPS units for navigation, and radio or laser communications.

Robot Arms

The next couple of items would be robot arms to give the propulsion unit the ability to do more complex tasks controlled from the ground. Items like robotic arms would be subject to a design trade-off. They would have to be made very rugged for gun delivery, versus a lighter weight version launched by conventional rockets. Arms could be made as segments with one or two joints, which are connected in series to make more flexible units, and have replaceable tool/manipulator ends for different tasks. The arms are designed as double-ended, so that either end can attach to a base or tool, and have a split joint to go from one shaft to two or more "fingers" or "arms".

Robot Attachments

These are tools that attach to the arms, and a rail car unit to move the arm from place to place.

Structural Base

This is a set of truss elements that can be assembled into larger arbitrary structures to which other parts of the growing assembly station will be attached. One approach is a **ball and stick** truss, with hubs at the intersections that have fittings at 90 and 45 degree angles. These are connected with struts of standard lengths to form the framework. The base truss might have a spacing of 1 meter, with adapters to scale up or down to other grid sizes as needed. Filler plates would span a truss bay to add rigidity or provide container spaces or additional mounting locations. The plates can be either perforated or solid as needed.

The basic structural system includes rails for moving robot arms and other items from place to place. The rails would extend a short distance from the hubs, with smooth joints to allow continuous motion. Either curved or pivoting sections would enable changing the plane of motion.

Utility Grid

The concept here is to have a redundant and modular utility system with different services (power, data, fuel lines) added as needed. One approach is to use a truss column as the utility carrier, and install support brackets to hold the various lines, with insulation or meteoroid blankets on the sides. That allows for easy access for additions or repair.

Power Units

There is another trade-off to do here for photovoltaic arrays, which are not suited to gun launch, but are lightweight, versus something like a Brayton generator, which in theory can be rugged. For low orbit, the power units would need some sort of storage, i.e. batteries, since sunlight is only available 60% of the time. To start with, simply attaching PV arrays to your structural base will provide a power supply

Electric Thruster Unit

For more extended missions that require more fuel, Ion or Plasma thrusters are added, which are more efficient than chemical thrusters.

Enclosed Modules

Some equipment, and humans, benefit from not being in vacuum. Other tasks benefit from temperature control, or keeping debris contained. For those sorts of requirements an enclosed module is needed. For early use, an inflatable module may be suitable. Finished modules are not suited to gun launch, but a fiber-reinforced aluminum tape could be launched as a spool, then formed around a mandrel to create larger shapes. Concentrated sunlight and pressure rollers can braze/solidify layers of tape until sufficient thickness is built up. That way the small cargo volume of the gun projectiles could be used to fabricate larger items. Once sufficient habitable volume and supplies are in place, then humans can start to work on the Assembly Station, but the initial construction will all be done via remote control.

Electric Propulsion

Electric propulsion typically has about ten times the fuel efficiency of chemical rockets. Thus they turn an exponential fuel requirement (fuel to push more fuel) into a nearly linear one for most Solar System missions. The timing of this step would be in parallel or soon after **Orbital Assembly** is started.

Near Term Electric Thruster Types

There are several kinds of electric thrusters that are good candidates for near term use. The selection here is based on state of development and usefulness:

- **49 Electrostatic Ion** - This knocks electrons of of gas atoms making them charged, which is called "ionized". Once charged, they can be accelerated by metal screens with a large voltage difference. Ion thrusters are used on some communications satellites, and the Dawn spacecraft currently exploring the asteroids Vesta and Ceres.
- **51 Microwave Heated Plasma** - This type uses microwave frequency heaters to heat the fuel. This is the same principle as a microwave oven, but much more intense. Above a certain temperature the heated atoms in the fuel will knock electrons of each other, turning it into a mixture of ions and electrons, which is called a **Plasma**. The plasma is contained and directed by magnetic fields. You need to do that because plasma is so hot it will melt anything it contacts, or cool itself down too much. In fact on Earth plasma is used as an efficient way to cut through metal. A version of this thruster is currently being developed on the ground, and will soon fly for testing on the Space Station. It's full name is **Variable Specific Impulse Magnetoplasma Rocket**, which is mercifully abbreviated to **VASIMR**. As a category they are called plasma thrusters.
- **72 Ionospheric Current** - This operates like a motor, running a current in a wire in a magnetic field. The return path for the current is the ionosphere. This method is limited to places with suitable magnetic field and ionosphere density, but low Earth orbit fortunately is such a location. The attraction is it does not require direct fuel use, only a little leaked plasma to make electrical contact with the ionosphere. The equivalent exhaust velocity as if it were a fuel-using engine is 250 km/s. Since low orbit is the first place we want to use, developing this type of thruster is a high priority. Note it is not as fully developed as the other types.

Electric vs Chemical Thrusters

All rockets work by tossing mass in one direction, and by Newton's Law (for every action, there is an equal and opposite reaction), the rest of the rocket gets pushed in the other direction. The faster you toss the mass, the more push (momentum) you get out of it. Conventional rockets burn fuel in a chamber then let it expand out a supersonic nozzle to get it going as fast as possible. The shape of the nozzle is governed by the physics of expanding gases, which is why they all look more or less the same. How fast it can get is limited by how hot the gas is and it's molecular weight. The best combination used today is burning Hydrogen and Oxygen in a ratio of 1:6 by weight. This produces mostly steam with a bit of Hydrogen left over to lower the average molecular weight. How fast the gas is going is technically called **exhaust velocity**, and is limited to about 4.5 km/s for this fuel type.

Electric thrusters are not limited by the energy produced by burning the fuel. They feed energy to the fuel from an external source, thus can get much higher exhaust velocity. This gives you more push from given amount of fuel. Since you have a finite amount of fuel to use this is more efficient in direct proportion to the increase in exhaust velocity. By analogy to automobiles, you are getting better "gas mileage".

The extremely high fuel efficiency is the key to why this type of thruster is important. If you are doing a lot of moving about in space the fuel savings outweigh (literally) the mass and cost of the power supply by a large margin. Conventional rockets only need a fairly lightweight fuel tank, but burn a lot more fuel. One drawback to electric thrusters when transporting humans is their relatively low thrust. This makes the trip times longer. There are various ways to work around that drawback. For example, passage through the Earth's radiation belts slowly would be unacceptable radiation exposure. So you can transport your main vehicle by electric thruster, taking weeks, and then deliver a crew in a small capsule taking hours once the main vehicle is outside the radiation belts.

Comparisons Between Types

All electric thruster types need an external power supply since the fuel is not self-heating as in chemical engines. The most common power supply used in space are photovoltaic panels. Those can get unwieldy at power levels of hundreds of kW or more, and their power output per area drops as the inverse square of distance from the Sun. So for some past and future missions a nuclear power source is preferred. Smaller size nuclear generators are based on isotope decay, and larger ones are full nuclear reactors. Any type of nuclear device brings both technical and political complications.

Electric thrusters cannot be used directly for launch or landing on large objects, because their thrust-to-mass ratio is significantly less than the local gravity acceleration. They can be used indirectly via space cable/elevator type systems. Chemical engines can reach vehicle thrust-to-mass ratios well above Earth gravity, which is one reason they have been the primary way to launch things to date.

Both Ion and Microwave Plasma thrusters have exhaust velocities in the range of 20 to 50 km/s, so are 4 to 10 times more fuel efficient than conventional rockets. Like electric devices on Earth, they are rated by how much power they use. The Dawn spacecraft has a 10 kW set of solar panels, and the VASIMR thruster in development is rated at 200 kW. Generally ion thrusters will maintain efficiency at lower power levels than plasma type thrusters because ion flow does not have to be restrained by containment fields, while plasma requires a field to keep it separated from the solid hardware. At small sizes the plasma volume vs total engine volume becomes small and efficiency drops.

For efficiency reasons, ion thrusters prefer high atomic weight fuels. The energy to ionize an atom is roughly constant across the Periodic Table, but does not contribute to thrust in this engine type. Thus using high weight fuels lowers the portion of total power used for ionization relative to acceleration. Typically Xenon is used as a fuel. Plasma thrusters can use most fuel types since their goal is to make the plasma extremely hot, on the order of a million degrees. By tuning the microwave generators, most atoms and molecules will absorb the energy. A key advantage of this is fuels like Oxygen or water are common in asteroids, so electric thrusters can be refueled locally, rather than having to bring all the fuel from Earth.

Electric Propulsion Applications

The following early missions can be performed starting with relatively small thrust levels, and working up to more ambitious missions.

Atmosphere Mining

This mission involves collecting air from the edge of the Earth's atmosphere for fuel and breathing. We start with a 50 kW solar array and a VASIMR type thruster which can generate 2 Newtons thrust at 40% efficiency and 20 km/s exhaust velocity. The solar arrays are assumed to use modern multi-layer cells with 30% efficiency, and have a power to mass ratio of 100 W/kg. The array will thus mass 500 kg, and we assume operates 30% of the time by intermittent use. The electric thruster can then produce an average thrust of 0.6 Newtons. At 200 km altitude, each square meter of collector generates 0.0129 Newtons of drag, so the total collector allowed area is 46 square meters to match the average thrust. This will collect 0.08 g/s, and the thruster consumes 0.03 g/s, leaving a net of 0.05 g/s. This amounts to 4.32 kg/day, or 3.15 times the solar array mass per year.

Later expansion would take the same thruster module to 200 kw power level and 5.7 N thrust at 50 km/s exhaust velocity and 60% operating time. The operating time is limited by the 40% of the orbit in the Earth's shadow. Thus average thrust is 3.42 N, and collection rate is 0.456 g/s. The thruster uses 0.114

g/s, leaving a net of 0.342 g/s.This is 29.5 kg/day or 10,785 kg/year, or 5.4 times the array mass per year. With a 15 year service life for the arrays, they can supply 75 times their mass in total. An electrodynamic thruster to make up for drag might improve on this even further

For human transport, where speed is important going through the radiation belt, the collected air can be separated for Oxygen, and mixed with added Hydrogen from Earth in a chemical thruster. Alternately a lower exhaust velocity, higher thrust electric thruster could be used, sacrificing fuel efficiency for fast transit. There are several plasma and arc jet thrusters that could do that job.

Orbital Cleanup and Maintenance

Earth orbit has accumulated debris from spacecraft explosions and collisions, and there are a number of non-functional satellites which only need a single part repaired or new fuel to function again. This mission involves using a range of electric thruster vehicle sizes to collect the debris, repair or refuel satellites on location, or bring them to the orbital platform for maintenance. Debris mass ranges down to centimeter or less in size, so it would be inefficient to send a large vehicle to collect it. Alternately, satellites can range up to several tons in mass. Therefore we select electric vehicle sizes to match the size of what is being collected or moved. For debris collection, several pieces in similar orbits can be collected in one trip to minimize fuel use and mission time. The fuel for these cleanup missions comes from the atmosphere mining. Depending on what the target objects are, we perform one or more of the following tasks:

- Collect orbital debris and either deliver it to a low enough orbit that it will decay and burn up quickly, or feed the debris into a processing unit to extract useful materials from.
- Return non-working satellite hardware to the orbital platform to be salvaged for working parts.
- Repair non-working satellites at the orbital platform with salvaged or new parts.
- Repair, refuel, or attach a new propulsion unit to existing satellite at their current location.
- Transport new cargo to higher orbits.

The tasks above are approximately in order of size and difficulty. Before salvaging used satellites, you would need to get permission from their original owners. The legal regime for broken debris pieces is unclear. If they are considered a menace to navigation, they might be removed without permission, or the original owners charged for cleanup.

Moved Text to be Merged

[MOVE to Phase 4B High Orbit] Further expansion of production may leads to power satellites, which beam energy to Earth for 24-hour power. If this can be done economically, it would likely be the largest export market to Earth. Solar-thermal with storage works in sunny climates on Earth, but many people don't live in such climates. Solar flux in space is 10 times higher than low sun climates on the ground. Energy delivered from orbit may prove cheaper overall, despite the extra cost of building in space. This is especially true if most of the materials for the satellites and their production equipment can be sourced from space, and the production is highly automated.

Section 4.6: Phase 4B - High Orbit Development

High Orbit Features

Region Definition

Our program defines the High Earth Orbit (HEO) region as orbits between 2700 km average altitude and the limit of the Earth's dominant gravitational influence, or **Hill Sphere**, at 1.5 million km. Where it would not be confused with high orbits around other bodies, we will often shorten the name to "High Orbit". Other sources describe this region as Medium Earth Orbit (2,000 to 35,000 km), High Earth Orbit (above 35,000), and **Synchronous Orbit** as the boundary between them. We are more concerned with the energy to reach the region and the local environment conditions for design purposes. They are similar enough to treat it as one region. Although it covers a large range of distances from Earth, it only represents the upper 25% of energy between the Earth's surface and escape. That's because gravity is an inverse square force and weakens rapidly as you increase distance. All of the region is in vacuum, and most of it is outside the Earth's magnetosphere and exposed to the solar wind. Solar energy levels in the region are also within 2% of that near Earth, so conditions in general are similar across the region.

This region completely surrounds the Low Orbit region (Phase 4A), and in turn is embedded in the Inner Interplanetary region (Phase 4C). The Moon is the most prominent object in in the area. It has it's own region embedded in the High Orbit one, with a radius of 35,000 km from the Moon's center and travels along with it. It also has its own program phase (5A) for development. Due to the Moon's gravity field, transport to and from the Lunar region requires additional energy, and local environment conditions are different, especially on the surface. Therefore for design purposes we assign it to a separate region.

Environment Parameters

High orbits are in sunlight 85-100% of the time, reaching the highest values when farther from the Earth and Moon. Temperature is determined mostly by the Sun and the cold Cosmic Background, but at the lower altitudes the Earth contributes a significant amount of reflected light and infrared heat. Orbit periods range from 2.5 hours to 7 months, so travel times by the most efficient routes can be long. Direct paths can be much faster, 12 days or less, at the expense of additional energy. Ping time varies from as little as 25 ms, which is not difficult, up to 10 seconds, which has a large impact on voice, real-time control, and electronic data. The upper part of the Earth's radiation belts, solar, and cosmic radiation create high to dangerous levels for people, without added shielding. Energy resources are abundant in this region, but material resources are low in their natural state. The Moon and Near Earth Asteroids can supply materials with fairly low transport energies.

Available Resources

Energy Resources

Material Resources

Industry Survey

The most popular satellite orbit, geostationary, at 35,000 km altitude, is in this region. This orbit has a period of 24 hours, which matches the Earth's rotation. Therefore satellites stay above a fixed ground location, and ground antennas can be stationary rather than having to track satellite motion. Synchronous orbit is in the outer fringe of the radiation belts, so manufacturing and human habitation tends to want to be higher up. Delivery, refueling, and maintenance of high orbit satellites is the main current market in this region. Likely the next step is supplying fuel and other supplies back to low orbit and for early interplanetary locations.Future industries are numerous, but depend on bringing costs down to affordable levels. This would happen incrementally as production for early markets bootstraps to larger levels. The total solar flux through this region is 500 million times what our civilization uses in 2015, and just the Moon can support a billion years of mining at the whole world's current rate. A small fraction of these resources can make our civilization sustainable for a very long time.

Project Drivers

Motivations

Economics

Technology

Placement

Development Projects

General Approach

Current and Near-Term Projects

Long-Term Projects

High Orbit Production

High orbits have abundant solar energy. It can be converted to electricity by solar panels or thermal generators, and used directly for heating using concentrating reflectors. Modern space solar panels and reflectors are very light weight relative to their power output, because they don't need to withstand gravity or weather. An initial stock of these power sources is enough to get early production going. Later expansions would be mostly self-built. The simplest product of all is radiation shielding for human crew. This only requires some crushing and sorting, then packaging into suitable containers around crew modules. Shielded modules allow extended crew stays in high orbit. The crews can operate a satellite maintenance and refueling station, and assist with early materials processing and production. To some extent the crew will be helped by remote control from Earth.

Next in difficulty are water and carbon compounds, from Carbonaceous-type asteroid. This requires 200-300C heat, which reflectors can supply, and a container and condenser to capture the vapors. Water and carbon can be chemically reformed to Oxygen and Hydrocarbons, which is a common high thrust rocket fuel. This is useful when transporting people through the radiation belts or for landing on the Moon. Water, carbon, air mined in low orbit, and possibly rock for soil can supply greenhouse modules, so that crews can produce their own food and recycling life support.

A higher temperature furnace can melt metallic asteroid pieces, add carbon to make steel, and then cast into basic shapes. With a supply of basic metal shapes, a seed factory that includes machine tools can then start making parts for additional machines. Basalt fibers made from Lunar basalts, and carbon fibers made from asteroid carbon compounds are very high strength. Other products would be vapor-deposited reflector sheets and parts for radiator panels. These are combined with high concentration solar cells from Earth to supply electricity at lower launch mass than complete panels. The same parts can be used to make furnaces and cooling systems for thermal processing of materials. Early production would therefore use a mix of pre-made processing equipment, like furnaces, and a growing set of equipment made on orbit. These will output a growing range of products, starting with fuel and other bulk supplies, and basic construction materials. Orbital industry can transition from importing modules and other station parts to building them locally, then exporting habitats to other destinations.

Ultimately large, comfortable space habitats can be built as permanent living space. These can be grown in layers, like an onion. Each layer adds a new compartmentalized pressure shell outside the previous ones. The outer few shells are in vacuum, and provide radiation, meteor impact, and thermal shielding. Inwards of that are pressurized areas with storage and mechanical equipment. Then comes living quarters and a central open space. As new layers are added, items are moved outwards to fill the larger space. Compared to building a large habitat all at once, this spreads the construction cost over time, and the habitat is only expanded when extra space is needed.

High Orbit Habitation

High Orbit Transport

Since high orbits are low in materials, they must be imported from elsewhere. The current method for transport from Earth uses a rocket to reach low orbit, then another rocket or electric propulsion to reach higher orbits. Electric propulsion is about ten times more fuel-efficient than chemical rockets, and is being used more in recent years. Electric thrusters require large amounts of solar power to operate, but efficient and lightweight solar panels have been developed in the last few decades (see **NREL Efficiency Chart**, Dec 2015 but frequently updated). However they are low thrust, and would expose unprotected humans to high radiation levels while slowly crossing the radiation belts. So some transport will have to be by alternate methods. Large amounts of propellant obtained in space relieves the penalty of doing this.

Bulk materials mined from Near Earth Asteroids are not time- or radiation-sensitive. They can be transported entirely by electric thrusters on tugs that make multiple trips. Since part of the product from these asteroids is more fuel for the tugs, the transport becomes self-sustaining once started. A tug can return about 750 times its hardware mass over a 15 year working life, while consuming about 17 times its mass in fuel over the same period. Tugs can also deliver hardware and finished products to other orbits as needed. The Moon is small enough that bulk materials can be tossed directly into orbit by an electric centrifuge. At 50% efficiency and 50% duty cycle from lunar night, a solar panel can power throwing 1000 times its own mass per year for 15 years. If the centrifuge is not too massive relative to the loads it throws, the overall mass return ratio is high. From low Lunar orbit, electric tugs take over and deliver the materials for processing. We want to source raw materials from both the Moon and Near Earth Asteroids, because they have different compositions.

High Orbit Services

Program Integration

Concept Details

Section 4.7: Phase 4C - Inner Interplanetary Development

The space between the major planets and moons has traditionally been viewed as devoid of use. They were to be gotten through as quickly as possible to reach the "real" destinations like the Moon or Mars. This view is incorrect and obsolete. Open space has always been known to have a large and constant flux of solar energy. At the Earth's distance from the Sun this is 1361 MW/km^2, or about one large nuclear plant, and there are 281 quadrillion square km around the Sun. From 1990 to late 2017, known **Near Earth Objects** alone in interplanetary space have grown from about 180 to 17,000, and we continue to **rapidly find more**. So there are plenty of available resources, and with new ideas on how to make use of them, widespread use of these regions should be possible.

The available energy flux, and quantity and type of materials varies significantly with distance from the Sun, as do local environment parameters. Therefore we divide development of interplanetary space into four phases (4C, 4D, 4E, and 4F) by distance, and discuss them separately. This one, Phase 4C - Inner Interplanetary Development, is closest both to the Sun and the Earth, so it comes first in sequence among the four. To date, only a few scientific missions have explored the region, and it hasn't been otherwise been used except for travel going elsewhere.

We begin concept exploration for this phase by describing the region's features in terms of the environment and available resources. We then survey industry categories to identify potential future activities in the region, and what will drive projects to implement these activities. These are combined into a development approach, and an initial list of projects by time and function. Later work will link the projects to each other and other parts of the program. Projects for which we have produced more details and calculations have that information included as the last major part of this section. An output of the concept exploration work is identifying what preparatory research and development is needed for this phase. This information is supplied to the earlier Phase 0I - R&D for Inner Interplanetary Development, since it must be completed before it can be used in this phase.

Inner Interplanetary Features

Environment Parameters

Available Resources

Energy Resources

Material Resources

Industry Survey

Project Drivers

Motivations

Economics

Technology

Placement

Development Projects

General Approach

Current and Near-Term Projects

Long-Term Projects

Inner Interplanetary Production

Inner Interplanetary Habitation

Inner Interplanetary T ransport

Inner Interplanetary Services

Program Integration

Concept Details

[Content to be merged]

7.3 - Phase 4C: Inner Interplanetary Locations

Inner Interplanetary Features

These locations are detached from the Earth's dominant gravity and orbit the Sun instead, although they may pass close to the Earth at times.They range from as close as equipment can function near the Sun to 1.8 AU, which is just beyond Mars' greatest distance from the Sun and where the Main Asteroid Belt starts. It excludes the four inner planets (Mercury, Venus, Earth, and Mars) and close orbits around them. Solar power is available 100% of the time in these orbits, but the intensity varies from 31% to many times that near Earth, depending on Solar distance. Ambient temperature correspondingly varies from very hot to 244K (-29C) for dark objects, less for bright or reflective ones. Travel time from Earth can range from months to years depending on orbit and propulsion method, and whether gravity assists from the planets are used. These save fuel, but usually require extra time. Solar and cosmic radiation are a moderately high background, with occasional flares/solar particle events that are much more intense, up to lethal human levels without shielding. Ping time ranges from a few seconds for orbits crossing near Earth, to over 45 minutes at 1.8 AU on the far side of the Sun from Earth, by way of a relay satellite. The Sun interrupts direct communication to theopposite side.

As noted in section 1.0, there are over 13,500 known asteroids closer than 1.3 AU, and several thousand more out to 1.8 AU. The largest is over 30 km in diameter, with about 100 times the mass of all the rock ever mined on Earth. So total material resources are very large. Asteroid orbits vary in size, are typically not circular, and somewhat tilted with respect to Earth's, so the energy required to reach a particular one varies. Timing matters also, since everything moves at different speeds in solar orbits. Efficient travel depends on your vehicle and the target being at the same place at the same time. The composition of asteroids vary across about a dozen spectral classes, indicating different chemical compositions. Only ten asteroids of the size we might mine have been visited by spacecraft. That does not count Vesta and Ceres in the Main Asteroid Belt. So most of our knowledge is telescopic and from examining meteorites that have fallen to Earth.

Economic Uses

There are not that many spacecraft currently in this region. They are mostly scientific probes in transit to other planets, or stationed at the Earth-Sun Lagrange points 1 and 2 (ESL1 and ESL-2). Future use is likely to start with asteroid mining and delivery to high Earth orbit with electric tugs. Most known asteroids in this region are too large to move as a whole. This is more because we can't find very small ones than because they don't exist. So mining would involve scraping material or grabbing a boulder from the surface of these larger objects. Prior to mining, a prospecting mission should visit multiple candidate asteroids and find out what they are made of in detail. As high Earth orbits become more developed, they can start to send equipment and seed factories to this region in addition to mining tugs. Since raw materials and full-time solar energy are available, the seed factories can grow into full scale factories and produce habitats, vehicles, and whatever else is needed.

Interplanetary Transport

The main transport system in this region is slow but efficient electric tugs. They can haul large loads of rock relative to their mass, up to 1000 tons for a 10 ton vehicle and 23 tons of fuel, but this depends on the orbit destination and velocity changes needed. They can also move crew habitats faster with a lighter load. Chemical rockets are used when fast velocity changes are needed, and solar sails may be effective in moving things even more slowly, but with no propellant use, once large lightweight reflectors can be made in orbit. Over time, a network of "transfer habitats" are built up. These are stationed in repeating orbits to particular destinations, and save having to move crew habitats each time for multiple trips. Centrifugal platforms can also be built over time to both provide comfortable gravity and fast velocity change with efficient propulsion. The platform mass serves as energy storage which can be transferred to payloads. Since the rotating structure can be relatively massive, it depends on large amounts of traffic to justify it economically, and production of high strength materials in orbit.

Interplanetary Production

Worldwide energy use on Earth is about 18 TeraWatts, and includes mining, processing, and manufacturing about 2 million kg/s of materials. Thus the energy intensity of Earth civilization is 9 MJ/kg on average. We will double this to allow for recycling of materials in space, and add 8 km/s of orbit velocity change, requiring 300 MJ of electric tug power. That covers a reasonable amount of interplanetary orbits. Transportation is thus the dominant energy use for new materials that need delivery. A space solar panel today produces 177 W/kg at the Earth's orbit, and produces the required 318 MJ in 20.8 days. Given an average life in space of 15 years, their total energy output is 260 times that needed to transport their own mass in raw materials and run the rest of civilization, including making replacement panels. Concentrating reflector and nuclear power sources are not yet developed enough for space to do calculate energy return ratios. They may turn out better or worse than solar panels, but as long as we have one known energy source with a high return ratio, we can base space industry on it.

The same process of bootstrapping production in high orbits can be used in interplanetary space. This starts with mining for export, then simple products made locally, and gradually bootstrapping to more complex ones. As distance increases, fewer raw materials would be brought from the Moon, and more from nearby asteroids. These asteroids are of different types, which provides a reasonable variety of materials to work with. If we restrict ourselves to within 20 degrees of the ecliptic plane, to maintain access to the planets and keep velocity changes lower, we have access to 1/3 of the Sun's total energy, or 1.3 x 10^26 Watts. This is 7 trillion times our current energy use, a number so large it is hard to imagine it could not sustain civilization.

With a network of systems developed around Earth and the Moon, the next step is to extend the network out to Mars and the Main Asteroid Belt. A similar process of setting up seed factories and adding facilities is followed as before.

System Concept

The system concept has the following main parts:

- Use free flying electric thruster powered vehicles to reach new locations that have useful resources.
- Set up seed factories in each location to build up industrial capacity, including more ships and seed factories for the next location.
- Produce fuel, life support supplies, and habitats at each location so they can be occupied permanently
- Build more Skyhooks to provide fast velocity changes, but get the benefit of electric propulsion efficiency.

High Earth Orbit Skyhook

In a previous step we defined a Low Earth Orbit (LEO) Skyhook with a tip velocity of 2400 m/s. Since its orbit velocity is 7474 m/s, at the top of its rotation it could release a cargo at a total of 9,874 m/s. At an altitude of 1,226 km, or a radius of 7,604 km from the Earth's center, with a Standard gravitational parameter of 398.6 x 10^12, we get an escape velocity of 10,239 m/s. Thus our LEO Skyhook is only 365 m/s short of reaching escape velocity. The LEO Skyhook can thus deliver cargo to an elliptical transfer orbit with a semi-major axis of the orbit of 54,278 km, and thus a high point of 100,952 km.

A circular orbit at that altitude has a velocity of 1,987 m/s, and the transfer orbit arrives at 743 m/s. The difference is 1,244 m/s. A tip velocity of 1,500 m/s or more would allow injection to Mars and Main Asteroid Belt transfer orbits, and any closer transfer orbits to the Moon and low inclination Near Earth Asteroids. A High Earth Orbit (HEO) Skyhook can therefore serve as a launch platform to any desired inner

Solar System destination orbit by selecting the radius, and thus velocity, and the time, which gives the direction, of release. This location is outside of the Earth's radiation belts, but it is also unprotected by the Earth's magnetosphere from solar and cosmic radiation. So human habitats will need radiation shielding. Such a high orbit is relatively easy to reach from the Moon or NEO's, so bulk matter for shielding will likely be brought from one of those sources.

The construction sequence would start with fetching Near Earth Orbit asteroid material and placing it in High orbit, and delivering equipment from Earth. Once a processing plant, factory, and habitat are set up, carbon from carbonaceous asteroids is used to make carbon fiber for the Skyhook. Initial velocity capacity would not be as high, and so more vehicle propulsion would be needed, but as the Skyhook grows, it can reach a wider range of orbits.Momentum changes are not free, so a substantial power supply and thruster set will be needed at the Skyhook. But since those do not have to be carried along with the vehicles who are getting their orbit changed, the propulsion can be as large and heavy as needed.

Inclination Stations

Near Earth Objects have a limited range of Solar orbit velocities in the ecliptic plane by definition. They also have a range of orbit inclinations, which results in a velocity component when crossing the ecliptic. The **Inclination Station** is a second Skyhook located in the vicinity of Earth, such as at one of the Lunar Lagrange points, and oriented perpendicular to the ecliptic plane. Therefore it can deliver cargo to and from inclined orbits to reach NEOs with less fuel and mission time, while the first HEO Skyhook operates in the ecliptic plane to reach Mars and the main Asteroid Belt. Additionally, flybys of planets can be used to further change orbit inclinations and reach other groups of asteroids. The Station itself will react to the average of cargo orbit changes. It will therefore need some propulsion to maintain position, but that will use less fuel than each vehicle doing it's own orbit maneuvers separately

Given sufficient traffic, it may make sense to have other Inclination Stations set up at different tilts to generate different combinations of ecliptic plane and inclination velocity changes. The velocity difference between the various high orbit Skyhooks will be very low, and release from less than their full radius will be sufficient to get between them. Since trip times between them will be short, the first one can have the bulk of the habitat and production facilities. The later ones can serve mostly as transit hubs, and not be as built up.

Transfer Habitats

Transfer orbits to NEOs or Mars will require trip times measured in months. For human passengers there is the risk of exposure to radiation, and also a need for food and life support. If you expect to make multiple trips, it makes sense to have habitats permanently in transfer orbits to those destinations. Then the mass of the shielding and greenhouses does not matter that much, as they are not moving once set up. The passengers would use a small vehicle between the HEO Skyhook and the Transfer Habitat when it passes near Earth, then ride in the habitat until it is near the destination, and then again use a small vehicle for arrival. Since only the passengers and cargo need to change velocity the total mass transferred per mission is greatly reduced.

All objects in the Solar System are in motion relative to each other, and transfer orbits only line up properly when, for example, the Earth and Mars are in the right relative positions. Thus a network of multiple Transfer Habitats in different orbits will be needed to deliver passengers and cargo to the right destinations at the right time. The source materials to build the Habitats would come mostly from whichever asteroids are already in the closest orbits. Depending on size, the NEO would be mined for materials, or if small, moved entire to the desired transfer orbit.

Optionally a transfer habitat would have a Skyhook attached to it to enable added delta-V for arriving or departing vehicles. This also provides an artificial gravity environment for the habitat. If that is not provided, then part of the habitat would be rotating to create artificial gravity. Whether to use a Skyhook or not will take more detailed analysis of the complete transportation network.

When the habitats are not ferrying passengers to Mars, they are exploiting **Near Habitat Asteroids**. Just like there is a set of asteroids whose orbits are "Near Earth", and so easy to reach for mining purposes, there will be a different set which are close to any given Transfer Habitat orbit. So you can busy yourself producing fuel, setting up manufacturing, and eventually have a space city there, just one that happens to get close to Mars periodically. When it does, you drop off humans and accumulated hardware at Phobos, where you are also building up facilities, and thence onward to Mars itself.

Orbit Characteristics

Since we are starting from Earth, we want the Habitat to pass by Earth on a regular basis. If we set the orbit period to be 1.50 years, it will do so every second orbit. The orbit will be an ellipse, and the long axis will be 2.62 AU. If the near point of the orbit is at Earth (1.00 AU), then the far point will be 1.62 AU, which is slightly past the average distance of Mars (1.52 AU). A velocity change of 3,340 m/s beyond Earth orbit is required to reach this orbit. This comes from a combination of the HEO Skyhook, Lunar gravity assist, propulsion on the transfer vehicle carrying crew and cargo between points, and possibly a Skyhook at the Transfer Habitat. The Habitat will align with Mars once every 7.5 years, and a velocity change of 4,440 m/s is needed at that end to match orbit. Again this would use a combination of propulsion, including Mars gravity assist. On early trips the transfer vehicle would need to do more work, but later a Mars orbit Skyhook can take up more of the velocity changes.

Since once per 7.5 years is not very often, you can place multiple habitats in a given orbit track, and use multiple orbit tracks spaced equally around the Earth's orbit at their near points. This will give more frequent opportunities to go back and forth from Earth to Mars. For 80% of their orbit cycle the habitats would would be doing mining and construction around the habitat, accessing asteroids in nearby orbits. The other 20% of the time they would also be carrying passengers and cargo for the trips to and from Mars, when they happen to line up. There may be a better arrangement of orbits and habitats to increase the fraction of time they can be used as a ferry service.

Section 4.7b - Orbital Mining

This step involves mining small asteroids in orbits close to the Earth for raw materials. These are called **Near Earth Objects** or **NEOs**. Although the Moon is physically closer, it has a significant gravity well, so it is not the best choice for a first mining location. Small asteroids have essentially no gravity well, so efficient **Electric Propulsion** can do the job of getting close, and then leaving. Both the Moon and NEO's have much smaller gravity wells than the Earth, so from an energy standpoint it is much easier to fetch materials from them. That advantage grows when the location for mining and eventual use of the materials are closer in energy terms. Once you start building installations on the Lunar surface, doing local mining there will be energy efficient, but that will be a later step.

Rationale for Orbital Mining

Past plans for space exploration or development have often selected the Moon or Mars as destinations. This has to do with them being large and obvious, and having lots of land area (we are a territorial species). That does not mean they are the correct early locations from an engineering standpoint. The Earth has a large gravity well that is difficult to climb out of, so it should be evident that you don't want to go down another deep gravity well right away, especially if you have to bring all your fuel from Earth. This question can be viewed in energy terms. The total energy to provide a kilogram of space hardware in orbit includes the energy to mine the raw materials, the energy to refine and fabricate the raw materials into finished parts, and the energy to transport all the matter involved to their final destinations. In the case of starting from Earth, the transport energy is by far the largest one, and increases the farther you go. Mining materials close to where you need them can reduce the transport energy, and thus lower total energy needed.

The Solar System is populated with a large number of small objects that do not have deep gravity wells. The ones in orbit around planets, like **Phobos** and **Deimos** are called **Satellites**. The ones not in orbit around a planet are called **Minor Planets**, and are named by location as **Asteroids** for the ones at or closer than Jupiter's orbit, and **Distant Minor Planets** for the ones beyond that. The major groups are further subdivided by orbit location. The ones which come close to Earth (NEOs) are what we are concerned with in this step, since the energy to reach them from Earth is low. There is not a great distinction in composition between smaller satellites and Minor Planets, just in where they happen to be located.

Even if the mission velocity to reach a particular NEO is somewhat higher than to reach the Lunar surface, and in many cases it is not, it can all be performed with electric propulsion. Electric thrusters are 6-10 times more fuel efficient than high thrust chemical rocket engines required for landing on the Moon. So total fuel use is less, often by a significant margin. Once Lunar infrastructure is in place, particularly to produce fuel, the Moon will be more accessible. But that will be covered in a later step. For now we start with what is easiest.

Steps Towards Mining

Currently (2012) we don't know enough about NEOs to properly plan mining them. So before actually doing the mining, we need to do prospecting to find out what is there in detail, and find out the best method of collecting the materials. A few missions (see **List of Asteroid Spacecraft** on Wikipedia) have gotten close enough to asteroids for detailed observations, but so far we have not visited specific ones that are easy to return materials from. Some NEOs are large enough to observe details from Earth (**NEO Physical Properties - Lupishko, 2002**), or are small but have passed close to the Earth, giving a short window for high resolution observation. The accumulated data so far are not sufficient, so additional observations and close up prospecting missions are needed.

Current Knowledge

Although we don't yet know enough about NEOs to mine them, we know a great deal more than nothing, and the data are accumulating rapidly.

Orbits:

The first things we discover about NEOs are generally their location in terms of orbital elements, and an estimate of size based on brightness. The survey telescopes to detect these types of objects are not the largest ones available, and they are working to their detection limits. Thus when an NEO is found, generally it is from single-pixel images against the stellar background at different times. The motion against the background allows calculating the orbit. The brightness of the pixel together with the distance from the orbit calculation lets us estimate size, based on an assumed surface color. At first we don't know the actual color of the NEO, and so the size remains uncertain until better data is collected.

Orbits of NEOs are grouped by size and shape into classes. All orbits are ellipses, and half the long axis, called the **semi-major axis** with symbol "a", are scaled to the Earth's, which is given the value of 1.00 astronomical units (AU). By definition NEO's have a < 1.3. In other words, their orbit is no more than 30% larger than the Earth's, or significantly closer than Mars is (a = 1.523 AU). Peri- and aphelion are closest and farthest points from the Sun respectively, in a given elliptical orbit. The major NEO classes by orbit size are mostly named after characteristic members of the class. These are:

- **Apollo** - a > 1.0 AU and crosses the Earth's orbit (perihelion < 1.02 AU)
- **Aten** - a < 1.0AU and crosses the Earth's orbit (aphelion < 1.0167 AU)
- **Amor** - Perihelion from 1.02 to 1.3 AU = always outside the Earth's orbit
- **Inner Earth Objects** - Aphelion < 0.983 AU = always inside the Earth's orbit

Sizes and Masses:

As of 10 Dec 2012, 9377 NEOs **have been discovered**, and about 900 new ones are being found per year. Their sizes can be approximated as a number N lager than diameter D (in km) by the formula

$$N = 1090 \times D^{-1.95}$$

or roughly the total number is proportional to the inverse square of size. Total mass of all the NEOs together is roughly 60 trillion tons, but this value is far from exact. The largest known is **1036 Ganymed**, which is about 32 km in diameter. A small number of NEO's have had density estimates made, which range about 1.5-2.67 g/cc. Comparing that to the most common meteorites, ordinary chondrites at 3.0-3.8 g/cc indicates 30-50% of their original volume is low density volatiles, which get burned off when reaching Earth, or is empty space. In the latter case, an NEO is better described as a pile of rocks than a solid object. One gram/cubic centimeter is also one metric ton per cubic meter, so the mass in tons can be found from the density times volume in cubic meters.

Rotation and Shape:

Rotation rates can be determined from variation in brightness, even when they only show up as a single pixel on a CCD detector. Therefore rotation rates of a considerable number of NEO's is known. They range from very low to 11 rotations per day, with larger numbers at the slower rates. About 60% are 5 rotations per day or less. The shape of the brightness variation curve gives an indication of the overall shape of the NEO.Without a detailed color map of the surface, we cannot distinguish variations due to shape from variations due to color (one part being lighter or darker than another part), but the indication is NEOs vary from round to about 3:1 maximum to minimum dimensions.

Composition:

The human eye is sensitive to three wavelengths of light, from which we get our sense of color and can determine a great deal about objects around us. Similarly, scientific instruments like large telescopes with spectrographs or color filters are sensitive to light of different wavelengths. A plot of brightness vs wavelength is called a **Spectrum**. Spectra from objects in space can be compared to those of meteorites and pure minerals. We can make a good guess at the composition of an NEO by this sort of comparison, given that enough light is collected to make the plot. An object too small and returning too little light for the available telescopes cannot be analyzed in this way, but it works for many of the lager NEOs, or small ones that happen to pass close to Earth.

Asteroid and meteorite composition is grouped into classes, and observed NEO spectra are about 62% S-complex class, 20% X-complex, 12% C-complex, and 6% other. The ones we observe are biased by the methods we use to observe them. A simple example is very dark objects are harder to spot in the first place, or to measure spectra from, so the actual composition mix is different. Obtaining spectra for the whole NEO population is ongoing (see Lazzarin et al, **SINEO: Spectrographic Investigation of Near Earth Objects**, 2004). As of 2008 around 2% had measured spectra. Observed spectra only tell us what the visible surface is like. The surface is affected by exposure to solar and cosmic radiation, and from impacts by other objects, so it is not a firm guide to the bulk properties of an NEO throughout it's volume.

The average life of an NEO is estimated to be 10 million years. They either crash onto one of the large bodies in the inner Solar System, or gravity effects change their orbit so they no longer fit the NEO category. Since the Solar System is 450 times older than this, there must be a constant source of new NEOs. This is mainly from the main Asteroid belt and extinct comets. Extinct comets are ones which have evaporated away all their volatile components and now are just the rocky remains. Comets and different parts of the main Asteroid belt formed at different distances from the Sun, thus at different temperatures. Therefore they collected different materials. Large enough asteroids became heated by radioactive decay and separated out into layers by density. Later collisions broke some of these up, and distributed pieces of different composition in different orbits. We expect the bulk composition of NEOs to generally vary within the limits set by the original Solar nebula from which they formed. However we cannot tell by their current orbits what that composition is because the orbits have been scrambled too much. We have to look at each individual object to determine what it is made of.

Morphology:

Morphology refers to the mechanical condition of a body. All objects in the Solar System that do not have atmospheres or surface renewal processes (crustal plates or vulcanism) are heavily cratered from random impacts of other objects. Close-up observations of asteroids, including NEOs, shows they are no different. Therefore their surface includes a **Regolith** (from the Latin meaning "blanket of stones"), a layer of impact debris and dust consisting of a mix of the original asteroid and whatever crashed into it. Over time, they may also have picked up dust and small rocks from the space they travel through via gravity or electrostatic forces, provided the material started out at low enough velocity to not just make more craters. High velocity impacts throw debris above the escape velocity of the object, thus adding to the population of smaller asteroids, rocks, and dust. From radar and thermal observations, the regolith layer is estimated to be on the order of a meter in thickness, with larger rocks sticking out, but this is highly variable by object and location on a given object. Internal structure of NEOs is so far poorly known, but they are expected to fall into three general classes. These classes have strong diferences in how they would be mined:

- Monolithic (single piece) fragments of larger objects, having strength typical of Earth rock.

- Rock piles of smaller objects held together by gravity, with no strength between the pieces, but strong pieces individually
- Very porous comet remains with essentially no strength.

Observing Programs

There are ongoing scientific observation programs directed at NEOs, so a mining program can build on that growing base of knowledge. The International Astronomical Union's **Minor Planet Center** has extensive data on known NEO's and current observing programs, from which most of the information in the previous section is derived. As of 10 Dec 2012, 859 NEOs larger than 1 km have been detected, which is estimated to be 90% of the total population in that size range. Discovery of these large NEOs is now less than 20/year. A further 100,000 are expected in the 100m to 1000m size range. Only about 8% of those have been found so far (see Large Synoptic Survey Telescope Science: **NEO Threat**). Even the smallest and least dense in that range (100m x 1.5 g/cc) would have a mass of about 800,000 tons, which is large compared to the largest objects placed in space (the ISS at about 450 tons), and a significant mass to mine. Thus the total NEO resource population is very large, and we have yet to even locate much of it yet. Beyond the NEO population, there are expected to be many millions of other minor planets in the Solar System, but accessing those is harder in energy terms, and is left for a later step.

At the current NEO discovery rate of 900 per year it would take about a century to find most of the remaining population larger than 100 meters. Since their orbits are randomly distributed, and so are the positions within their orbits, opportunities for exploration or mining missions increase about linearly with the number of known objects. This is currently about 10% per year. If the cost is not too great it would be worth increasing the number and size of telescopes searching for NEOs to increase the discovery rate. Looking from Earth, we only see the smaller ones if they get close enough to show up on telescope instruments. It may be worthwhile to send dedicated spacecraft to different parts of the NEO orbit range, as proposed by the **B612 Foundation**. It would be especially useful to search the orbits easiest to reach from Earth up close, to more easily find those particular ones. Even though only a fraction of all NEOs have been found, a good number are easy to reach (for example see Elvis et al, **Ultra-Low Delta-v Objects**, section 9, 2011).

There will be other NEOs whose orbits are very elliptical or inclined, which will not be good early mining candidates but will be an Earth impact hazard. For those type, a different search strategy is needed. A full sky search will find both types, but one looking for especially easy to reach NEOs will be looking in specific parts of the sky near the plane of the Earth's orbit. There is nothing preventing a telescope from doing both, it's just a matter of what part of it's observing time is used for each type of search. The recently started **Large Synoptic Survey Telescope** project, which is much larger than past asteroid search telescopes, can efficiently do both mining and hazard surveys.

As noted above, only a few NEOs have been observed with large enough telescopes to get their spectra, and thus a start at determining what they are made of. A dedicated observing program with larger telescopes will be needed to fill in the spectral data for a large percentage of NEOs, compared to the 1-2 m diameter class telescopes have been used so far for discovery. The LSST mentioned previously is sufficiently large for this purpose, as are other existing and planned telescopes, but observing time on large telescopes is in high demand. The LSST is one of these large telescopes, and only 15% of it's time may be dedicated to asteroid searches. So a dedicated or semi-dedicated telescope for gathering spectra would be very useful if you wanted to reach a certain percentage coverage by a specific date.

Prospector Missions

Only a limited amount of composition data can be gathered by ground-based telescopes. For more detailed information, getting the instruments closer is essential. These can be called prospecting missions in the mining sense. There are several types:

Meteorites:

These are former NEOs which got so near they conveniently crashed into the Earth. This makes them easy to collect relative to objects still in space. They are direct samples of the asteroid population, but they are not unmodified samples. Entry through the Earth's atmosphere, impact, and weathering (if they have been on the surface for a long time) have all changed them from their pre-impact state. Still, a great deal of useful knowledge can be gained from them because we can apply all the available scientific instruments to examine them. Projects to gather meteorites contribute to the general knowledge pool for future mining, and are inexpensive relative to space missions. Even more helpful is tracking incoming meteorites with telescopes and radar, and then finding them on the ground, since we can then associate them with a specific source orbit. The mix of material types for meteorites is different than NEOs because re-entry and weathering affects some types more than others, but we can adjust for this once we know enough.

Remote Missions:

To get better data on NEOs in their current locations we must go to the source. The size and weight of spacecraft instruments is severely limited compared to what we have available on Earth, therefore we want to get as close as possible to use their limited sensitivity. For NEOs that would be to fly nearby, go into close orbit, or land on the object. The ongoing **Dawn** mission to the largest Main Belt asteroids, (4) Vesta and (1) Ceres, is an example of this mission type. Dawn goes into orbit around the object, and observes it remotely with cameras and other instruments, then sends the data to Earth. Close observation of a reasonable sample of typical objects could lead to extrapolation of other object's characteristics based on measurements from Earth, particularly their spectrum in various wavelengths. A single spacecraft with sufficient fuel and electric thrusters could visit multiple objects, like Dawn is doing.

Sample Return Missions:

A more ambitious mission type would collect a sample or samples from one or more NEOs, and return them to Earth for analysis. This could be either directly via a re-entry capsule, or indirectly by delivering it to the Space Station, after which the samples are sent down to Earth. Sample returns allow using the full range of Earth instruments. One spacecraft, depending on design, could return multiple samples from a single object, samples from different objects, or fly multiple missions to different destinations, with re-fueling at Earth each time. For direct sample missions the spacecraft either has to land, send an impactor to throw a cloud of material to be collected from orbit, or use some kind of scoop or mining bucket from a distance. Each approach would need a different design. Landing on a small object is not like landing on a large body like the Moon. The gravity levels are so low that staying in place may require anchoring. Otherwise just using a robotic arm to scoop up a soil sample might lever the rest of the lander off the surface.

Candidate Selection

We need to select which asteroids to investigate, and then later mine. Our initial criteria will be based on how easy they are to reach, size, and composition. Mission velocity determines how much fuel will be used per trip, so lower is better. Size determines total amount of mass that can be mined, so larger is better. Composition of NEOs varies, so preference on that basis depends what we need the materials for.

Velocity Class (km/s)	Known NEOs
3.8 to 4.0	9
4.0 to 4.5	96
4.5 to 5.0	303
5.0 to 5.5	450
5.5 to 6.0	729
6.0 to 7.0	2464
7.0 to 15.0	4611
15.0 to 26.1	322

Mission Velocity

Starting with the first criterion, The table on the right lists the number of known NEOs by velocity range from low Earth orbit, based on the **JPL Table of Low Velocity Asteroids**. The table presented here is as of mid-2012. The JPL source data will change over time as new NEOs are discovered. Note that 5.9 km/s is the velocity required to reach the Lunar surface from Low Earth Orbit, purely in velocity terms. So on that basis many of these are easier to reach than the Moon.

Asteroid missions can be done entirely with efficient electric thrusters. So even ones that require higher actual velocity can be done with less fuel than landing on the Moon. To find out how much less, we assume only the Low Lunar Orbit to Lunar Surface part (1.9 km/s) has to be done with high thrust chemical rockets. They are required for this portion because once you go below orbital velocity you will impact if you do not land quickly. In contrast electric thrusters need a 41% higher total mission velocity, because they use continuous spiral thrusting paths rather than the short burns of chemical rockets have. Their exhaust velocity is 50 km/s rather than the 4.5 km/s at best for chemical, requiring 11.1 times less fuel per velocity increment. The net advantage after considering thrust profile is 7.85 times less fuel per velocity increment for electric. The first 4 km/s are assumed to be the same in both cases, using electric propulsion to go from Low Earth Orbit to Low Lunar Orbit or Earth escape. If we multiply the Lunar landing part by 7.85 we get 14.9 km/s as the additional velocity we can take for an asteroid mission and use the same total fuel as landing on the Moon. We add this to the 4 km/s in both cases to get 18.9 km/s total mission velocity. From the table we can see that very few, in fact only 87 out of 8986, or less than 1% of NEOs require more fuel to reach than the Moon given these assumptions.

The best asteroid candidates need about 11.5% fuel mass to final mass, while Lunar landings need 70.75%. So under the best circumstances, asteroids are 6 times easier to reach in fuel terms. If we look at the low velocity NEOs that require less than 5.3 km/s ideal velocity to reach, the fuel required is under 16.2% and the advantage is at least 4.3 times over a Lunar landing. Additional data on low velocity NEOs can be found in Elvis et. al. (referenced above), although that paper neglects electric propulsion.

Size and Mass

Albedo	Range	High to Low		Mass (Megatons, A=0.25
0.50	0.25	0.05	H (abs mag)	D=1300, Spherical)
470	670	1500	18.0	205
370	530	1200	18.5	101
300	420	940	19.0	50
240	330	740	19.5	24.5
190	260	590	20.0	12.0
150	210	470	20.5	6.3
120	170	370	21.0	3.3
95	130	300	21.5	1.5
75	110	240	22.0	0.9

The size of most NEOs is poorly determined by telescope observation from Earth because their image is less than 1 pixel in the attached camera.For now we usually estimate size from brightness. This is still somewhat uncertain because brightness is the product of physical size and reflectivity (albedo). Most NEOs are various shades of dark gray to black, and the darker the color, the less light reaches the telescope. Until we can make more detailed observations, we use a range of estimated albedo to produce a size estimate. To further make our mass estimates uncertain, we usually do not know the detailed shapes or density. The latter can range from about 1300 to 7800 kg/m^3 depending on how solid the object is and what it is made of. We shall assume a candidate NEO should be at least 1 million tons in mass to be worth exploring in detail and setting up mining. If the unknown albedo is assumed to be at the higher end of the range, at a given brightness it will be smaller. If we also assume the lower end of the density range we can make a minimum mass estimate based on brightness, as shown in the last column of the table on the right.

Current Candidate NEOs

The following list of candidate NEOs is drawn from the same JPL table noted above. The selection criteria are:

- Electric propulsion delta-v from LEO < 7.5 km/s. The table values are for high thrust delta-v, which is lower by a factor of 1.414 (square root of 2), but the ratio is constant, so can be used for selection. Actual orbital mining missions will likely start from a high orbit, so the actual mission velocities will be lower. The difference will be a constant value representing the delta-V from LEO to your actual starting point, so the relative order of candidates wi stay the same.

- Absolute magnitude (H) < 22.0. This gives a probable size of 110-240 meters and a probably mass greater than 900,000 tons.

Composition has not been used to narrow the selection because not many of the asteroids have had their spectra taken and none have been visited. Therefore this list only represents the state of knowledge as of 2012. As additional objects are discovered, more details about known ones accumulated, and needs for particular materials to extract are developed, the best candidates will change. The columns are:

- Provisional name - This is year of discovery and a serial number within the year
- Delta-v in km/s - As noted, the values are for high thrust missions.
- Relative Delta-v - These are the previous delta-V value compared to velocities to reach the Lunar and Martian surfaces, assuming aerobraking at Mars.
- H - Absolute visual magnitude. This is the brightness of the object at a standard viewing distance of 1 AU and fully lit by the Sun. The actual brightness from Earth varies constantly as their positions change.
- a - Semi-major axis in AU. This is half the long axis of the elliptical orbit of the object.
- e - Eccentricity. This is the (difference in closest and farthest distance from the Sun)/(sum of closest and farthest distance = major axis). It is a measure of the shape of the orbit and ranges from 0 for a circular orbit to 1 for a parabolic orbit which just reaches solar escape velocity
- i - Inclination in degrees. This is the tilt of the orbit plane with reference to the Earth's orbit.
- Notes - The permanent object number and name are noted if they have been assigned. They are not immediately assigned on discovery because multiple observers might detect the same object, it may be human-made (which have their own numbering), or it may not be in a permanent orbit. 1999 RQ36 is the planned target of the Osiris-Rex mission.

Provisional	Delta-V	Relative-	-Velocity	Brightness	Orbit Axis	Eccentricity	Inclination	Notes
Name	(km/s)	(to Moon)	(to Mars)	(H in magnitude)	(a in AU)	(e)	(degrees)	
2011 CG2	4.125	0.688	0.655	21.5	1.177	0.159	2.8	
2001 US16	4.428	0.738	0.703	20.2	1.356	0.253	1.9	(89136)
2002 NV16	4.456	0.743	0.707	21.3	1.238	0.220	3.5	
1993 BX3	4.500	0.750	0.714	20.9	1.395	0.281	2.8	(65717)
2003 GA	4.511	0.752	0.716	21.1	1.282	0.191	3.8	
2000 FJ10	4.560	0.760	0.724	20.9	1.319	0.234	5.3	(190491)
1998 HG49	4.615	0.769	0.732	21.8	1.201	0.113	4.2	(251732)
2004 KE1	4.619	0.770	0.733	21.6	1.299	0.181	2.9	
1998 SF36	4.632	0.772	0.735	19.2	1.324	0.280	1.6	(25143) Itokawa
1999 JU3	4.646	0.774	0.737	19.2	1.190	0.190	5.9	(162173)
1997 WB21	4.672	0.779	0.742	20.3	1.461	0.317	3.4	
1994 CJ1	4.698	0.783	0.746	21.4	1.489	0.325	2.3	
2006 SU49	4.711	0.785	0.748	19.5	1.413	0.312	2.5	(292220)
2012 DK6	4.735	0.789	0.752	21.0	1.243	0.166	6.3	
2003 CC	4.743	0.791	0.753	20.3	1.500	0.327	2.3	
2008 WN2	4.752	0.792	0.754	20.8	1.418	0.312	3.7	
2012 DK61	4.735	0.789	0.752	21.0	1.243	0.166	6.3	
2003 CC	4.743	0.791	0.753	20.3	1.500	0.327	2.3	
2008 WN2	4.752	0.792	0.754	20.8	1.418	0.312	3.7	
2000 YJ11	4.767	0.794	0.757	20.7	1.313	0.232	7.3	(162783)
2008 YS27	4.774	0.796	0.758	21.1	1.468	0.317	4.9	
2008 DG5	4.785	0.798	0.760	19.7	1.256	0.243	5.7	
2001 WC47	4.794	0.799	0.761	18.9	1.399	0.242	2.9	(141018)
2008 SO	4.827	0.804	0.766	20.7	1.331	0.234	7.1	
2009 SC15	4.830	0.805	0.767	21.6	1.265	0.179	6.8	
2002 SR	4.852	0.809	0.770	21.6	1.179	0.196	6.7	
2009 SQ104	4.873	0.812	0.773	20.9	1.284	0.279	4.0	
2000 EA14	4.876	0.813	0.774	21.0	1.117	0.203	3.6	
2002 TC70	4.886	0.814	0.776	20.9	1.369	0.197	2.1	(253062)
1989 ML	4.888	0.815	0.776	19.3	1.272	0.136	4.4	(10302)
1996 FO3	4.901	0.817	0.778	20.5	1.443	0.290	5.8	
2008 TD2	4.923	0.820	0.781	21.7	1.530	0.334	4.0	
2011 AK5	4.940	0.823	0.784	21.5	1.188	0.230	5.5	
2010 TH19	4.960	0.827	0.787	20.5	1.464	0.310	6.8	
2011 BT15	4.971	0.829	0.789	21.7	1.297	0.304	1.7	
2001 QC34	4.972	0.829	0.789	20.0	1.128	0.187	6.2	

2006 UQ17	4.972	0.829	0.789	21.9	1.624	0.381	1.7	
2003 GY	4.973	0.829	0.789	20.1	1.380	0.317	4.7	
1982 DB	4.979	0.830	0.790	18.2	1.489	0.360	1.4	(4660) Nereus
2006 YF	4.987	0.831	0.792	20.9	1.109	0.199	4.7	
2004 PJ2	5.009	0.835	0.795	21.4	1.418	0.342	2.6	
2011 AM24	5.012	0.835	0.796	20.4	1.178	0.150	9.1	
1999 NA5	5.032	0.839	0.799	20.4	1.436	0.249	4.3	(264308)
2011 GD60	5.034	0.839	0.799	21.7	1.083	0.162	6.1	
2001 VB76	5.034	0.839	0.799	20.4	1.459	0.348	4.2	
2010 PR10	5.051	0.842	0.802	21.7	1.198	0.176	9.2	
2009 DL46	5.069	0.845	0.805	21.6	1.456	0.305	7.9	
2000 SL10	5.081	0.847	0.807	21.9	1.372	0.339	1.5	
1999 RQ36	5.087	0.848	0.808	20.9	1.126	0.204	6.0	(101955) Osiris-Rex mission
2007 CN26	5.089	0.848	0.808	20.8	1.295	0.270	7.6	
2011 UW158	5.093	0.849	0.808	19.4	1.617	0.375	4.6	
1996 GT	5.098	0.850	0.809	18.0	1.644	0.384	3.4	(65803) Didymos
2004 BE86	5.107	0.851	0.811	20.9	1.441	0.237	3.8	
1999 ND43	5.131	0.855	0.814	19.1	1.523	0.314	5.6	(36017)
1999 YR14	5.133	0.856	0.815	18.9	1.654	0.401	3.7	
2012 EY11	5.135	0.856	0.815	21.9	1.148	0.151	9.0	
2009 DN1	5.136	0.856	0.815	20.3	1.442	0.286	7.9	
2000 LY27	5.136	0.856	0.815	17.0	1.309	0.213	9.0	(67367)
2008 HS3	5.138	0.856	0.816	21.7	1.351	0.226	8.2	
2001 XP88	5.155	0.859	0.818	20.6	1.347	0.194	6.7	
1994 CN2	5.159	0.860	0.819	16.8	1.573	0.395	1.4	(136618)
2000 QK130	5.187	0.865	0.823	20.6	1.181	0.262	4.7	(216985)
2009 EK1	5.188	0.865	0.823	21.4	1.242	0.230	9.1	
2000 WO148	5.192	0.865	0.824	20.7	1.642	0.376	4.4	
2005 JS108	5.197	0.866	0.825	19.2	1.356	0.322	6.0	(187040)
2007 HX3	5.204	0.867	0.826	20.0	1.527	0.312	6.1	
2011 EM51	5.213	0.869	0.828	21.9	1.321	0.335	1.9	
2007 BF72	5.229	0.871	0.830	19.7	1.433	0.215	4.1	(311925)
1997 WT22	5.247	0.875	0.833	18.8	1.486	0.306	8.2	(136839)
2001 QQ142	5.249	0.875	0.833	18.4	1.423	0.311	9.3	(139622)
2001 SW169	5.250	0.875	0.833	19.0	1.248	0.052	3.6	(163000)

2005 QA5	5.274	0.879	0.837	21.2	1.390	0.211	6.8	
1994 UG	5.276	0.879	0.837	21.0	1.238	0.293	5.2	
2002 LJ3	5.283	0.881	0.839	18.3	1.462	0.275	7.6	(99799)
2004 JA27	5.296	0.883	0.841	19.4	1.666	0.423	2.3	(164211)
1993 HA	5.302	0.884	0.842	20.1	1.278	0.144	7.7	(52381)

Continue to page 2

Section 4.7b - Orbital Mining (page 2)

Return to page 1

NEO Ore Types

An ore is a raw material whose composition is of economic value. On Earth, we generally class as ores those source rocks which are particularly high in one component. In space, the current cost of almost everything is thousands of times higher than on Earth, so almost any kind of mass has economic value. Raw unprocessed rock can be used for radiation shielding, so even that has considerable value. So we consider all NEOs to be ores, but they still have differing compositions which makes them suited for different uses. As noted on Page 1, we don't yet have very good data on the bulk composition and distribution of materials for most Near Earth Objects, but we can make a list of general categories based on observations of the larger main belt asteroids, meteorites, and the small fraction of NEOs with observed spectra. The main categories in the widely used Tholen **Asteroid Spectral Types** are further subdivided into distinct subtypes.The main types are:

C-group

These represent 75% of asteroids in general, and an undetermined but probably similar fraction of NEOs. These have a similar composition to the Sun and the original Solar nebula from which the Solar System formed, minus hydrogen, helium, and some volatile compounds. The missing components are gases at the temperature and pressure of NEOs, and so have evaporated away. Unlike the Gas Giant planets, NEOs have no gravity well to speak of to trap these light components. The spectra of these objects are similar to the **Carbonaceous Chondrite** type meteorites, which are assumed to come from the same source bodies. Specific types within the C-group are enriched or depleted of specific components. Typically they consist of a physical mix of grains of different composition, including a significant iron-nickel component (3–20% depending on subtype), mineral silicates and oxides, sulfides, water (up to 22%), and organic (carbon) compounds from which the group gets its name. C-group objects are very dark, with an **albedo** (reflectivity in sunlight) of 0.03 to 0.10, so they are hard to spot initially. Initial size estimates tend to underestimate their size, since those are based on average albedos, which are assumed to be 0.10.

S-type

This represents about 17% of asteroids in general, and get their name after their **Stony** makeup. They contain mostly iron and magnesium silicates, but without the carbon in the C type. They originate from the inner Main Belt asteroids, where they are dominant at <2.2 AU, common at <3 AU, and rare farther out.

X-Group

These include the high metallic fraction asteroids, ranging from 50 to 100% metals, with the remainder as stony inclusions. Some of them are nickel-iron alloy, thought to come from the metallic core of larger asteroids which were later destroyed. The remainder are thought to have crystallized without separating in bulk from smaller asteroid bodies or being exposed to high temperatures. Therefore the metallic parts have diferent compositions.

Other Types

There are a number of other groups, the A, D, T, Q, R, and V-types, which are present in small numbers, but do not form a coherent large group by composition.

Example Mining Missions

There are approximately 600,000 known minor planets of all types throughout the Solar System. **Near Earth Objects** (NEOs) are the subset of about 9,300 known minor planets which approach within 1.3 times the Earth's distance from the Sun (1.3 AU). This is substantially closer than Mars' orbit (1.52 AU) and the Main Belt asteroids (mostly 2.1-3.3 AU), and therefore easier to reach from Earth.

Newly discovered minor planets get identified by year and a serial number within the year, such as **2011 AG5**. When their orbits are better determined, they get a permanent serial number and possibly a name, such as **(1) Ceres** or **(4) Vesta**. A 2011 study by Landau and Strange at the Jet Propulsion Laboratory (**Near Earth Asteroids Accessible to Human Exploration** AAS 11-446) tabulates a number of such with particularly easy to reach orbits. Their estimated masses range from 50 tons to 6 billion tons (for (175706) 1996 FG3) so in most cases bringing back the entire object is not practical for an early mining mission. Instead, we will look at several mining mission types with diferent difficulties and time frames when you would consider doing them:

- Mining the Earth's Debris Belt
- Return an entire small (~500 ton) NEO,
- Gather 1000 tons of unprocessed regolith from the surface of a larger object,
- Extracting fuel onsite and pushing the entire object towards Earth, and
- Extracting finished materials and only returning those to Earth.

Debris Belt Mining

This would both be a demonstration of mining techniques, such as capturing a non-cooperating target, and also be useful by removing hazards from Earth orbit. The **Debris Belt** is the accumulated defunct spacecraft and pieces of spacecraft from accidental explosions and collisions. By combining atmosphere mining for fuel, and electric propulsion, sufficient velocity capacity would be available to rendezvous with the scattered objects and either return them to an Assembly Platform for recycling, or to a low enough orbit where drag will quickly cause them to re-enter. For efficiency, multiple objects in similar orbits would be collected on a single trip, and several collector ships of different sizes would be used. It would not be efficient to send a large ship after a small object. The debris belt is currently about ten times the hazard of natural meteoroids, so cleaning up this trash would greatly reduce the hazards to functioning spacecraft, for which the operators (or their insurance companies) should be willing to pay.

Some non-functional spacecraft only have a single failed part, or ran out of fuel, and could be made to operate again. In that case they can repaired and placed back in service. Others could be salvaged for parts, or their materials recycled. Anything not useable would be disposed of by re-entry Such a salvage and repair business could pay for itself.

Small NEO Return Mission

Purpose - Return an entire small asteroid to near Earth as a demonstration, for scientific examination, and for prototyping processing methods.

Description - The Keck Institute for Space Studies performed a study in 2011-2012 in some detail of returning an entire small NEO about 7 meters in diameter and 250-1300 tons in mass (see **Asteroid Retrieval Feasibility Study** or **Alternate Source**). The NEO would be bagged in it's entirety by an inflatable structure that is then cinched down for transport. Due to the low gravity of such a small object, this method will prevent loss of any dust and rocks from its surface. Small objects are hard to characterize from the Earth, which accounts for the large range in mass estimate. If a small flyby or orbiter mission is sent ahead of time, the mass and other characteristics can be better determined. This will likely be required before doing detailed design of the mining tug.

Assumptions - As for any technical study the assumptions made affect the results. Major assumptions include:

- The starting point is Low Earth Orbit and the destination is high Lunar orbit for the returned NEO, after which it would be examined scientifically and processed for materials. Lunar gravity assist is used in both directions in the mission.

This mission requires a total velocity of 9.56 km/s for a particular NEO selected to model the design around.

- It is performed as a complete mission with a single launch on an Atlas V 551-class launcher and a single spacecraft. This limits the total spacecraft mass at launch, including having all the propellant pre-loaded. It also limits the solar array size to what can be deployed automatically from a single payload volume.
- The propulsion system uses 40 kW net at end of life (plus 1.2 kW for other electrical needs), and has an estimated mass not counting propellant tanks of 1,370 kg, thus a power/mass ratio of 33.25 kg/kW. The power system is 60% efficient, and assumes the solar arrays lose 20% of their original efficiency traveling outward from the Earth's radiation belts. It uses 5 x 10 kW ion thrusters with an exhaust velocity of 30 km/s. The Xenon propellant has a tank mass fraction of 4.25% above the propellant mass.

Results - Overall spacecraft start mass is estimated at 14.7 to 18.8 tons for returning 250 to 1300 tons, thus providing a 17 to 69 NEO return mass/spacecraft mass ratio. Fuel consumed is from to 8.8 to 12.9 tons, giving an NEO mass/fuel mass ratio of 28 to 101. This mission was designed as a single return mission, but if sufficient propellant can be extracted from an NEO, ranging from 1-4% of its mass, future missions can be self-sustaining on fuel. With 40 kW power for propulsion, the total mission time ranges from 4.0 to 10.2 years depending on asteroid mass. The NEO mass return rate is then 62.5 to 127.5 tons per year. If higher return rates are needed, larger solar panels and higher power thrusters would be needed, scaling approximately linearly with mass return rate. At some power level solar panels will get unreasonably large and you would need multiple spacecraft. Current Space Station array size, if replaced with modern cells, would be around 300 kW, and a practical limit might be 2 MW for a truss with multiple arrays and thrusters. Large arrays would break the single launch with no orbital assembly assumption. For higher power levels than what is practical for solar, a nuclear reactor would be the likely power source.

Bulk Regolith Return Mission

Purpose - Return a maximum of 1000 tons of bulk NEO surface material (regolith) per trip to feed a processing plant in High Earth Orbit (HEO).

Description - HEO is from 10 Earth radii (above the radiation belts) to the limits of the Earth-Moon system. Because gravity decreases as the inverse square of distance, all HEO locations have similar velocity requirements and are near Earth escape. An HEO location minimizes the total round-trip velocity for each asteroid mission. The Mining Tug is assumed to make repeat trips using part of the extracted propellant from the previous trip to refuel with. Asteroid orbits close to the Earth's are low energy, but also low difference in orbit period. Therefore they come close to the Earth at long intervals, and are not good candidates for successive trips. Thus different destination asteroids are chosen for each trip based on orbital opportunities. Major parts of the Mining Tug, like the solar arrays and thrusters, are repaired or replaced as needed at the processing plant to keep the Tug functioning. The tug is therefore designed for orbital assembly and refueling, both to reduce initial component launch masses from Earth and to enable maintenance/upgrades.

Assumptions

- A radiation-hardened solar array is used to transit through the Van Allen radiation belts to High Earth Orbit. The radiation protection uses cover glass over the solar cells, which absorbs the radiation before it can damage the active layers in the cells. The cover glass adds mass and slightly lowers efficiency by absorbing and reflecting some of the sunlight. Lower mass and more efficient arrays are used for transit from the processing plant to the NEO's and back. The high efficiency arrays are folded and protected by radiation shielding during delivery to HEO. Unprotected arrays can lose 20% of their efficiency passing through the radiation belt under their own power, so using a hardened array for that job makes a significant difference.
- The Small NEO Return Mission captured an entire small NEO with minimal disturbance to preserve science. This mission assumes gathering 1000 tons of surface rocks and dust from a larger NEO for easier processing once returned. Larger NEOs are easier to detect and determine properties from Earth, thus giving more mission candidates. This will lower the mission velocities slightly by having more to choose from, and allow a better schedule for later missions.
- Higher exhaust velocity is used to save propellant, and higher power is used for faster transit times. For a production mining operation efficiency and annual mass delivered are important. Propellant is assumed to be Oxygen since that can be extracted in large quantities in all but metallic type NEOs. This makes the mining operation self-sustaining on fuel. We assume the Ad Astra VASIMR plasma thruster and solar arrays have a combined mass of 10 kg/kW, exhaust velocity of 50 km/s, and Oxygen propellant tank mass fraction of 10%. (source: Ad Astra Rocket Company, **A Survey of Missions using VASIMR for Flexible Space Exploration**, 2010).

- The first trip may fetch a smaller load, such as 200 tons, to reduce the original fuel mass required from Earth, and make the second trip sooner to bootstrap fuel productionAs much as half this mass is Oxygen of varying extraction difficulty, which is more than enough to fuel full loads on later trips.

Calculations

Trajectory:

For mission planning in general, we work backwards from the destination to find out what we need to start with. We assume you can use Lunar Gravity Assist in both directions to help depart and return to Earth orbit. This mission does not need to return to Low Earth Orbit (LEO), but some final products will. Aerobrake passes through the upper atmosphere can be used for smaller cargo returns to LEO and save fuel. There are several detailed trajectory simulation programs to calculate exact trajectories to a known destination and back. For now, we don't know which NEO we are going to or when. In fact, it may not even be discovered yet. For now propulsion will be calculated based on generic propulsive velocity required of 4 km/s outbound and 1 km/s inbound.

About 1% of known NEOs have an ideal velocity to reach from LEO of 4.5 km/s or less (see previous page). It is assumed that future discoveries will maintain the proportion of low velocity candidates. Of those candidates, a subset will have suitable size, composition, and orbit timing. Starting from HEO (> 64,000 km radius), velocity to reach Lunar flyby will be about 1.2 km/s, and Solar orbit change to reach the NEO will be about 2.8 km/s. We choose an NEO which will make a close pass to the Earth at the time we want to return from it. That way we can make a relatively small velocity change to set up an Earth or Moon gravity assist to help with the return trip. Since the mass will be much larger on the return leg, optimizing for this part reduces total propellant required. The return velocity is then estimated at 1 km/s.

Propellant:

A 200 kW plasma thruster is in development, so if we have 5 of them plus a 1 MW solar array, the estimated hardware mass is 10 tons. The plasma thruster has an exhaust velocity of 50 km/s, so from the rocket equation we can calculate the mass ratio is 1.02 to produce 1 km/s velocity change on the return trip. Mass values will be given for both 200 and 1000 tons returned respectively in parentheses as (200,1000) tons. Given an end mass of (210,1010) tons (vehicle hardware plus returned regolith), the fuel for this part of the mission is (4.24,20.4) tons. On the outgoing leg of the mission we assume 4 km/s is required. The mass ratio is thus 1.083 against an empty vehicle mass of 10 tons, plus (4.24,20.4) tons fuel for the return trip, so (1.19,2.5) tons of fuel are needed for this part. Thus total fuel needed is (5.43, 22.9) tons.

The overall return ratio is 200 or 1000 tons regolith vs 5.43 or 22.9 tons fuel, or 37 to 43:1. This is a very attractive ratio as long as we can extract a reasonable fraction of the regolith mass as useful products. In particular, if we can extract at least 23 tons of Oxygen, or 11.5% of the first trip's returned mass, the later full mining missions become self-sustaining on fuel, and the return ratio over the life of the mining vehicle goes up dramatically. If the vehicle can make 5-10 trips before major hardware replacement, the net mass returned after fuel used will be 4100 to 9000 tons against a launch mass of 15.5 tons, for an overall return ratio of 264 to 580.

The particular numbers above will change according to which NEO you select, and which start and end dates are used for a mission. At present the known collection of NEOs is growing by 10% a year, and this is expected to accelerate as larger telescopes come on line, so there are more to choose from over time. NEOs are always moving in their respective orbits, so their distance from Earth constantly changes, and thus so does the mission pathEach thruster uses 9.85 kg of fuel per day. With 5 thrusters that becomes 49.25 kg/day at full power. From the total amount of fuel used we can calculate the engine run times as 110 and 465 days, and the total trip will be that plus whatever coasting time is needed due to orbit positions, and time at the NEO to do the mining. Very roughly we allow 200 days coast time and 100 days mining time, and thus a total trip time of 1.125 to 2.1 years.

The original assumption of 1000 tons returned is not a fixed requirement. Within reason that can be larger or smaller, as long as the main components of the mining ship scale linearly. Plasma thrusters are not as efficient at lower power levels, so below about 80 kW it will make sense to use ion thrusters. For reliability, one or two spare engines should be added above the number needed for propulsion. Large solar arrays by their nature have enough dplicated parts to be reliable.

Large NEO Return Mission

'*Purpose* - Return a whole larger NEO to Earth Orbit for later mining. This would be second-generation mining with substantial markets and infrastructure required.

Description - We will use the example of **2011 AG5**, an NEO expected to pass within 300,000 km of the Earth in 2040. The objective is to shift it's orbit enough to do a gravity flyby in 2040, to set up for capture into the Earth-Moon system later. Once captured, it is then mined for materials. Given an estimated mass of 4 million tons, electric thrusters using solar arrays do not seem feasible. Several approaches are possible:

- Use a powerful nuclear reactor or solar concentrating thermal-electric generator to supply power to larger plasma thrusters. The fuel comes from the asteroid itself, which requires an extraction plant on-site. Such a large operation might require setting up a habitat and crew on the asteroid to operate and control it. For a 500 m/s propulsive change, 40,000 tons of fuel would be required. If we allow 5 years to make the velocity change, then 8,000 tons a year are needed.
- Put a container around the NEO so material is not lost, add or build a pusher plate/shock absorber unit made from asteroid material, and use one or more small nuclear devices to make the velocity change. The energy to change the asteroid velocity by 500 m/s is 125 kJ/kg, thus the total energy needed is 500 TeraJoules. This is equivalent to 120 kiloTons of TNT in energy, plus an efficiency factor. If the pusher plate is 25% efficient in capturing the explosive energy, then 500 kT of devices are needed, divided into however many units are needed to match the shock absorber and container strength. Use of nuclear devices presents obvious hazards, as does any asteroid material return when the impact energy from objects approaching Earth is 15 times their mass in TNT.

Product Return Mission

Purpose - Set up a processing plant on the NEO, and only return finished items rather than bulk regolith.

Description - The farther from Earth the final user is, the more it makes sense to process the NEO material from a nearby asteroid, rather than returning to Earth orbit first, and then send out again to the final location. For multiple destinations, this would imply multiple processing plants, so they would need to be smaller and more efficient to operate than a centralized plant somewhere near Earth/

Mining Tug Design

(this section is preliminary)

Small asteroids are typically rotating. So the mining concept is to enter synchronous orbit around the body, and using some method like sending down a scoop on a cable to haul materials up to the tug. This avoids the issues of trying to land on a moving target, and the relatively low thrust-to-mass of a loaded tug, which might have difficulty getting off even a small asteroid. Another reason to avoid landing is losing power from being on the night side of the asteroid. In theory landing at a pole of the asteroid simplifies the moving target problem, but that restricts your choice of mining locations.

Sufficiently small objects would not have much rotation velocity, but have the problem of staying attached if you land. We do not know enough about surface cohesion yet to design anchor systems. It may be necessary to run cables entirely around the object in order to stay anchored.

Section 4.7c - Spaceport Network

Space Elevators have been a theoretical transportation method since 1895. The original idea is impractical to build. This step adds a much more practical design as a transport hub for getting from one orbit to another quickly and efficiently. Initial construction can use materials from Earth, but in larger sizes or locations beyond Earth orbit using local materials is assumed.

Skyhook Concept

The popular concept of a space elevator is based on the original design proposed by Tsiolkovsky in the late 19th century. It involves a single tower/cable extending all the way past Geosynchronous (24 hour) Earth Orbit (GEO). If the center of mass is at GEO and matches the Earth's daily rotation it will appear to hang motionless relative to the ground. Getting to space in theory then becomes an elevator ride. There are several problems with this simplistic design:

- The depth of the Earth's gravity well (6378 g-km) exceeds the scale length of the best available materials (350 km for carbon fiber) by 18 times, which then requires a structure-to-payload mass ratio of 65 million to 1. This would require more carbon fiber than the world makes to lift a reasonable cargo mass, and would never be economical because it would take too long to transport sufficient payload to justify the massive cost of that much structure.

- It is of no use for delivering cargo to low orbits. Release points somewhat below GEO result in elliptical orbits with a low perigee, but lower circular orbits cannot be reached. It also is of no use transporting cargo from the ground when partially built.

- Even with a magnetically levitated elevator car running at 300 km/hr, it will still take 5 days to deliver one payload to GEO, and you can only deliver one payload at a time.

- A single cable catastrophically fails when hit by natural or man-made debris. A cable over 35,000 km long has a lot of area exposed to such hits.

The **Skyhook** concept addresses all these problems. Instead of a static cable that stays over a fixed location, it can be either a rotating cable in a moving orbit, much like two spokes of an imaginary wheel rolling around the Equator, or a non-rotating cable in a moving orbit that maintains a vertical orientation relative to the parent body

- The non-rotating orbiting Skyhook is a much shorter version of the planetary surface to geostationary orbit Space Elevator that does not reach down to the surface of the parent body, is much lighter in mass, can be affordably built with existing materials and technology, and in its mature form, is cost competitive with what is thought to be realistically achievable using a Space Elevator. It works by starting from a relatively low altitude orbit and hanging a cable down to just above the Earth's atmosphere. Since the lower end of the cable is moving at less than orbital velocity for its altitude, a launch vehicle flying to the bottom of the Skyhook can carry a larger payload then it could carry on its own. When the cable is long enough, Single Stage to Skyhook flight with a reusable launch vehicle becomes possible at a price that is affordable to just about anyone.

- A full orbit velocity rotating Skyhook reduces the structural requirement to about 2868 g-km, because only the tip sees the full Earth's gravity. The center is in orbit and thus has zero acceleration load. This immediately reduces the theoretical mass ratio from 65 million to 3,620:1. There is still an exponential relation of mass to tip velocity. Since conventional rockets also have an exponential relation of mass to velocity, it makes sense to split the job between both, because the sum of two exponents, for example e^2 + e^2 = 14.8 is less than a single exponent of the combined powers, i.e e^4 = 54.6. The optimum division of work between the Skyhook and vehicle coming from Earth will depend on technical details and costs, but a simple division of half to each results in a theoretical Skyhook mass ratio of 60:1. A real design will be heavier, but 60:1 is a feasible starting point, where 65 million is not. Reaching half of orbit velocity for a single stage rocket with a life of many flights is quite feasible.

- Assuming the tip is at 1 gravity, a rotating Skyhook with a tip velocity of 30-50% of orbit velocity has a radius of 500-1400 km. The center point needs to be that altitude plus enough that the tip does not dip into thick atmosphere and create drag (100-200 km). Releasing from the center of the Skyhook at 600-1600 km altitude allows access to low Earth orbits.

- A partially built Skyhook can still function because the remainder of the velocity is provided by the Earth vehicle. During construction the velocity split is more towards the Earth vehicle. This reduces payload mass, but it can still deliver some. In particular, if part of your payload is more Skyhook structure, that payload pays for itself in increased payload on later trips. This is a literal version of "lifting yourself by your bootstraps".
- The same fast elevator car at 300 km/hr can reach the center in 2 to 5.3 hours. If your destination is high orbit or Earth escape, you do not have to ride the elevator at all. You wait half a rotation of the Skyhook and let go, at which point you are going at orbit velocity plus tip velocity. To imagine this, think of the top point of a bicycle wheel. It moves faster than the center relative to the ground. A half rotation takes only 10 to 20 minutes.
- Space debris cannot be eliminated. Even if all the man-made junk in Earth orbit is eliminated, the natural flux of meteors will continue. Therefore the Skyhook design has to take that into account. The most practical way to do that is to use multiple redundant cables to distribute the load such that cutting one or two is not catastrophic. The cables should be spaced far enough apart that any single object will only hit one or two strands. The strands should also be cross-connected periodically to distribute the load around a break. Repairing a break then becomes replacing a short segment of one strand. Since you have the ability to install segments during original construction, you are able to replace segments as a maintenance job.

Skyhook Applications

The Moon and Mars have smaller gravity wells than the Earth, by ratio of 22 and 5 respectively, so Skyhooks with the same materials can do more of the transportation job relative to Earth. But in this step-by-step combined system example, getting off the Earth comes first. We will discuss the other locations here, but the actual construction will be delayed until easy transport to those locations is needed.

Lunar Skyhooks

There are two systems that are feasible because of the small gravity well of the Moon. The first is a catapult system to launch bulk materials off the Moon. The catapult uses a rotating cable driven by an electric motor to throw payloads directly into Lunar orbit, where they are picked up by a collector system. The second is an orbiting Skyhook which can deposit and pick up cargoes at zero velocity close to the surface.

Catapults - Basalt fibers are similar to fiberglass in that they are an extruded mineral. They have a strength of 4800 MPa, or 80% of carbon fiber, and a density of 2.7 g/cc, or 50% higher than than carbon. Thus the scale length of Basalt fiber is 178 km, or about half that of carbon fiber. The dark areas (Maria or seas) of the Moon are covered in basalt lava[1], so there is an a very large supply of raw materials. If a Lunar catapult delivers basalt to an orbital processing factory, or spools of fiber already spun on the Moon, it would be possible to build a Lunar Skyhook out of local materials. The choice of Lunar basalt for a Skyhook would have to be compared to the higher performance carbon fiber brought from Earth or made from NEO carbon. Certainly for Lunar surface construction it would have the advantage of being very local.

Catapults could also be built on the Earth or Mars, but for Earth it would need to be placed above the atmosphere on a tall tower to get significant velocity. It is probably not the best method when compared to the alternatives. Mars is much smaller, has less atmosphere, and very tall volcanoes that a catapult can be placed on. So it is worth considering placing a catapult there to deliver materials to orbit. Any catapult (Lunar or other) will need a significant power supply for the motor. In order to not waste the rotation energy when stopping to load the next cargo, it makes sense to consider two catapults, and use one as a generator to supply power to the other. Motor-generator efficiencies can be above 90%, so most of the energy could be recycled.

Skyhook - The Moon's gravity well is equal to 287 km at 1.0 Earth gravity. Thus even for the lower strength basalt fibers, the gravity well is only 1.6 times the scale length of 178 km. For the higher strength carbon fibers with 361 km scale length the ratio is 0.80. Note that scale length is based on breaking strength, actual designs will use lower loads and have overhead above a bare cable. For Earth the theoretical gravity well to carbon fiber scale length ratio is 18, so it is much easier to build a Lunar Skyhook on a relative basis. Another way to say this is the Lunar orbit velocity of 1680 m/s is less than the 2400 m/s tip velocity assumed below for the Earth Skyhook. That provides all of the velocity between the Lunar surface and orbit, while the Earth orbit version only provides 1/3. Since escape velocity is 1.414 times circular orbit velocity, and a full Lunar Skyhook is capable of releasing cargo at 2.0 times orbital velocity, it can handle cargo well beyond escape velocity. By climbing to a chosen radius from the center, and timing when you let go, you can get a wide variety of orbits.

Besides using it as transport to and from the Moon, there is also an opportunity for work crews to use the Skyhook at 1.0 gravity as a rest location, because we don't know the long term effects of 0.16 gravity on the human body. A lunar surface alternative is to use centrifuges to get 1.0 gravity or whatever level is needed. A full Lunar Skyhook would have a radius of 283 km and a rotation period of 17 min 40 sec at 1.0 gravity. Since Lunar orbit period is 108 minutes or longer depending on altitude, the Skyhook will make 6 or more rotations per orbit. If the orbit is equatorial, that allows it to service multiple locations around the Lunar equator, and transport cargo between those points at orbital speed at no cost.

It is an open question if an equatorial orbit is best. A polar orbit would let the Skyhook reach any point on the Lunar surface, but generally only twice a month. The Moon rotates very slowly, so the benefit of the rotation towards the orbit velocity is only 4.6 m/s, 1% of the Earth's contribution. A polar orbit can be arranged as a **Sun-synchronous orbit**, where the orbit plane always is in sunlight, while an equatorial orbit is in shadow about 40% of the time. Thus the solar arrays that power the Skyhook are more effective in the polar orbit. You can have both Skyhooks in orbit around the Moon, as long as you arrange their orbits to never intersect, such as by using different altitudes. In that case you might want to make the g-forces at the tips higher so the radius is smaller, and move humans quickly up the cable to a more comfortable g-level.

In any Lunar Skyhook, a lander vehicle will need some propulsion because the Moon is not a perfect sphere. So the tips need to stay high enough to miss any high points of the terrain, and some maneuvering is needed for an accurate landing. If you have two Skyhooks at different altitudes, the vehicle will need more fuel to land and take of.

Asteroid Skyhooks

The largest asteroid, 1 Ceres, is 487 km in radius at the equator, with a day length (rotation period) of 9.074 hours. Therefore the equator is moving at 94 m/s. Orbit velocity is estimated at 360 m/s. The exact number will be found when the Dawn spacecraft arrives at Ceres in 2015 (it is in orbit around the 2nd largest asteroid, Vesta, as of 2012). A Skyhook thus needs only the difference of 266 m/s in order to land and pick up cargo and then toss them at more than escape velocity. The radius in this case at 1 g works out to 7.25 km. This is small enough that it could be built near the Earth, and then transported whole to Ceres. Setting it up in orbit would allow mining of the largest asteroid with easy access. A synchronous space elevator would be longer and not provide a 1 gravity environment, but could be used to launch cargo into transfer orbits away from Ceres or capture incoming cargo.

For small asteroids, a Skyhook isn’t necessary for surface access. Even low efficiency chemical rockets do not use much fuel to land, and you can just mechanically throw stuff into orbit or escape.

Mars Skyhooks

Pavonis Mons is one of the large mountains on Mars. Since it is located on the Equator, it is an ideal location for some kind of transport system. Candidates include a centrifuge launcher like with the Moon, or a linear accelerator. The higher mass of Mars makes it more difficult than for the Moon, but a ground-based transport system can still do most or all of the job of reaching orbit velocity. Similarly, Mars orbit velocity of 3.6 km/s is within reach of a Skyhook, and there are two convenient former asteroids (Phobos and Deimos) as a source of building materials.A Martian Skyhook would likely be placed in low Mars orbit, with the ability to transfer down to the surface and up to Phobos, Deimos, or escape orbits.

Split Systems

A catapult can be used in combination with a Skyhook to enable higher velocity missions with lower total mass ratios. Bodies as small as the Moon do not require very large mass ratios o reach orbit, so doing a split system will not gain much at the cost of the extra complexity. Conversely the Earth has a fairly dense atmosphere, so a high velocity centrifuge would see a lot of drag unless placed on a very tall structure. The best location for a split system turns out to be Mars, particularly with its tall mountains that are in near vacuum at their peaks. By dividing the velocity between two systems, it becomes 1.8 km/s each, which can be reached with existing materials and conservative mass ratios.

Design Parameters

The Earth orbit Skyhook does not have a fixed design as noted above, but rather grows over time. We also do not as yet know what an optimum size will be for given circumstances. A concrete example, however, lets us examine the feasibility and understand what is needed for the various parts.We will assume a tip velocity of 2400 m/s for this example, or roughly 1/3 of orbit velocity, and derive the other characteristics.

Design Inputs

Tip Velocity = 2400 m/s

Tip Acceleration = 10 m/s^2 - Earth surface gravity is 9.80665 m/s^2. We use 10 for simplicity. That provides normal gravity for any humans on the Skyhook.

Derived Values

Skyhook Radius = 576 km - This is found by solving the centrifugal acceleration formula (a = v^2/r) for the radius.

Rotation Period = 25 minutes - We know the circumference of a circle by 2 x pi x r, or 3619 km in this case.Dividing by tip velocity gives the time. For convenience to reach the Skyhook from a launch site, the numbers can be adjusted so the period is an even fraction of the orbit time, i.e. 100 minute orbit with 25 minute rotation time. That way the landing platform will be in the same place each time relative to the launch site.

Orbit Altitude = 750 km - If the tips of the Skyhook reach deeply into the Earth's atmosphere, that will cause drag and heating, and eventually cause the Skyhook to fall down. By placing the tips at least 175 km altitude, then the center must be that plus the radius high. The exact height will be a trade of between less drag, and ease to reach from the ground.

Orbit Velocity = 7474 m/s - Found from the formula below where G is the Gravitational constant, M is the mass of the Earth, and r i the orbit radius, which is the Earth's radius plus the orbit altitude:

$$v_o \approx \sqrt{\frac{GM}{r}}$$

Launch Vehicle Payload = 13% - A good chemical rocket would have an exhaust velocity of 4.5 km/s and empty weight of 10%. Without a Skyhook, the total velocity required is about 9 km/s, which results in a payload fraction of 3.5%. Subtracting the 2.4 km/s provided by the Skyhook results in a payload of 13%, or 3.7 times higher. The exact numbers will vary depending on the launch vehicle design, but that gives an idea of the payload improvement the Skyhook can provide, and thus part of the reason to build it. The greatest advantage of a Skyhook is not from the increased payload it provides, but using some of the increase to increase the fatigue life of the vehicle, which is highly non-linear - typically ten times higher for a 10% addition in structure. Airplanes and rockets cost about the same per kg to build. This is not surprising since both are built by aerospace companies out of the same materials. The vast difference in transport cost is due to airplanes flying about 20,000 times during their service life, and rockets usually only flying once. By taking some of the payload increase from a Skyhook and applying it to giving the launch vehicle a long operating life, the operating cost will be vastly reduced.

Payback Time = 1 to 76 launches (average of 43) - If we remove the last 100 m/s of tip velocity, the launch vehicle payload falls to 12.56% from 13.07%. So the incremental benefit of the last 100 m/s is 0.51% of the vehicle mass. Assume we use Torayca T1000G carbon fiber as our main cable material.It has a tensile strength of 6370 MPa and a density of 1.8 g/cc. We allow 40% overhead mass above the bare fiber for a finished cable system, and a 2.0 factor of safety. Thus the working strength is reduced to 2275 MPa, and at a tip acceleration of 10 m/s, the working length becomes 126.4 km.

The 2400 m/s Skyhook has a radius of 576 km, and the acceleration varies from 0 at the center to 10 m/s at the tip, so effective length is half, or 288 km.The cable mass ratio is then e ^ (288/126.4) = 9.762:1Subtracting the payload mass gives a theoretical cable mass of 8.762. Since this calculation is only for one arm of the Skyhook, we double it to 17.524. The payload of the Skyhook is the launch vehicle payload + the empty vehicle structure without fuel (10% of launch weight), for a total of 23.07% of launch weight. The

Skyhook cable mass is then 404.3% of the launch vehicle weight. Doing the same calculations for the 2300 m/s Skyhook, we have a radius of 529 km, effective length of 264.5 km, mass ratio of 8.106, and cable mass of 16.211 times the vehicle arrival mass of 22.56% = 365.7%. The incremental cable mass is then 404.3 - 365.7 = 38.6%. Since we gain 0.51% in payload from this incremental addition to the Skyhook, it pays for itself in payload mass in 76 launches if we use the launch vehicle to deliver the added cable. If we get the extra cable from another source, such as our hypervelocity gun, or from NEO carbon, the payback could be much faster.

Doing the same type of calculation over the whole Skyhook, we have a gain in payload from 3.5 to 13%, or 9.5% of liftoff mass. The Skyhook cable mass is 404.3% of liftoff mass, so the payback time in increased payload is 43 launches. The first 300 m/s increment of the Skyhook increases payload from 3.53% to 4.46%, a gain of 0.93% of liftoff mass. It would have a radius of 9 km, and a cable mass of 7.25% times arrival mass, or 0.98% of launch mass. Thus the payback time is 1 launch, and the cable fits in roughly 1/4 of a payload on it's delivery flight. So the first part of the Skyhook has an immediate payback and is very desirable. Note that a Skyhook massing less than the arrival mass would not have enough orbital energy to give the vehicle on arrival. It would need to be attached to a larger "ballast" mass such as an assembly platform, bulk mined materials, or collected space debris.

Mass payback is not the same as cost payback, but if we assume they are for now, at a relatively low rate of one launch/month, payback takes about 3.6 years, which is reasonable from an economic standpoint. Since the cable mass grows exponentially with tip velocity, the earlier parts will pay back faster, and growth beyond this point will take longer at a fixed launch rate. A real payback analysis will have to take into account real cost instead of mass ratios, and actual launch rates. If less expensive launch systems from Earth are developed, then the cost benefit of a Skyhook goes down, even if the payload mass increase stays the same. On the other hand, if traffic rates go up, do does the cost benefit. No matter what numbers are used, the exponential growth of the Skyhook mass with tip velocity will eventually limit it's size for economic reasons. That is when the incremental cost of making the Skyhook larger becomes more than the incremental payload increase is worth. We emphasize that limit can change over time, however, as new materials become available, different methods of cable delivery are used, and traffic rates change. The Skyhook/launch vehicle interaction is a good example of why combined systems have to be looked at in their entirety, and not as single technologies or methods.

Design Components

Structure

The main structural component of the Skyhook will be the tension strands. In addition there will be secondary structure holding the strands in position, and for the landing platform, propulsion system, habitats, and other items attached to the structure. Cables are not a stable structure unless they are in tension, so for the central portion of the Skyhook before it is set rotating, we assume a core rigid truss structure. The initial total radius is 2500 m, counting core plus cables. This allows 1.0 gravity acceleration at the tips with a rotation period of 100 seconds. The latter number is chosen so that humans are not dis-oriented by rapid rotation. Initial habitats would be placed at the 2500 m radius. Strands are installed both lengthwise and in parallel to expand the Skyhook.

Damage Tolerant Design - Man-made satellites and orbital debris are the largest impact hazard for the structure. For now we assume no cleanup of Earth orbit, though that is desirable. Doing so would reduce the risks by about ten times. If we assume there is a 0.5 mm protective sheath around the strand core, then objects smaller than half that thickness will merely make a crater and not penetrate the core. If we assume a large Skyhook supports 1000 tons of payload at 1.0 gravity, with an 8:1 mass ratio, we have a total load of 10 MN. Given the 2275 MPa working strength, that requires a total area of 44 cm^2 of cable. Assume each strand is 2 cm in diameter. Then we need 14 load carrying strands, and some number of extra strands for damage tolerance, which we will use 7 for now, giving 21 total strands.

Any debris object larger than 1/3 the strand diameter will likely cause enough damage that it fails. Based on a 1995 orbital debris assessment [2] there were about 1 million objects that size and larger. That produces an impact flux of about 10 ^ -4 per square meter per year. If we want a 1% chance of strand failure per year, then we are allowed 100 square meters of area. Since the diameter is assumed to be 2 cm, then the allowed length is 5000 meters. We place cross-connecting rings at that interval to distribute load around a failed strand. At each ring, there are 7 points where 3 strands each fan out. Since by design two are required to handle the load, the failure of the third by debris impact does not cause any reduction in the Skyhook's total load carrying ability.

We do not want a single object to impact too many strands at one time. In the worst case, the largest object in orbit besides the Skyhook, currently the Space Station, should not hit more than our reserve of 7 extra strands. In reality, 99% of damaging debris is smaller than 30 cm, and the Space Station is under active control, so it should not ever hit the Skyhook, but we are looking at a worst case. Since the Station is about 120m wide, if the strands are arranged in a circle, and there are 21 total, a 120m object should not intersect more than 120 degrees of the perimeter. Doing a little geometry yields 120m = 50% of the radius, thus the diameter of the circle is about 480 meters, and the 7 attachment points are spaced about 200 meters apart. A truss spans between each attach point, making a 7-sided ring. If a given strand is damaged, then it is simply replaced by the same construction method the Skyhook was built in the first place. As long as strands can be replaced at least as fast as they are damaged, the Skyhook can be maintained indefinitely.

The above calculations are an example. For a real design, you would find the optimum strand diameter and count, rather than just assume 2 cm and 21.The real debris population is not all in orbits that could intersect the Skyhook. For example, the Space Station is about 400 km altitude, and could only intersect with the bottom 225 km of the Skyhook when it is in the vertical position. Some efforts may be made to clean up orbital debris. But even if not, we can make a reasonable design that can withstand worst case damage and reduce expected damage to a 1% per year maintenance job.

Landing Platform - This functions somewhat like the deck of an aircraft carrier, in that it is a mobile platform which vehicles land and take off from. We assumed above that the Skyhook structure supports 1000 tons of load. This includes everything besides the main structure, including the landing platform. Arriving vehicles would have a smaller mass that is added to the load temporarily. Unlike zero-g docking, which is done slowly, the landing platform is rotating at 1 gravity, so landings will be similar speed to landing on Earth. The size of the platform will be governed by the accuracy of the vehicle navigation. The design can be a horizontal platform, or something like a latching hook or arresting cable, such as used on aircraft carriers. In that case the Skyhook name becomes a literal description. An alternate method is have a vertical capture net. It's as wide and tall as needed to make a good target. The vehicle has redundant capture latches deployed ahead of it, and arrives slightly faster so it runs into the net, and snags multiple cables.

It is assumed the landing will be automated for uncrewed cargo delivery, with radar, lidar, and other aids to getting within the landing target area. The landing platform is made a multiple of the navigation accuracy in size to have a high probability of hitting the target. The best design is an open question, but since landing at 1 gravity has been solved multiple times on Earth, it should be solvable for this task.

After delivering it's payload, the vehicle just needs to let go or be pushed off the platform at the right time. At it's low point, the platform is sub-orbital, so the vehicle will automatically re-enter. The vehicle will be moving at 4,600 m/s relative to the Equator, compared to 7,400 m/s for a rocket without a Skyhook trying to re-enter. Therefore the vehicle has to dissipate 39% as much kinetic energy, which makes the heat shield design much easier.

Low Gravity Platforms- Low- or Zero-gravity is desirable for some tasks in Space. You can place platforms or pressurized habitats at chosen distances from the Skyhook center and get any value between 0 and 1 gravity. For true zero-g, you would need to de-spin the structure so it does not rotate along with the rest of the Skyhook. Likely this will be a structure extending along the axis of rotation like the axle of a wheel, where the Skyhook cables would be spokes of the wheel.

Electric Propulsion

Due to fundamental conservation of energy, transporting more payloads up than down causes the Skyhook to lose orbital energy and eventually re-enter if not corrected. We use the highly efficient electric thrusters developed in an earlier step to maintain orbit. In effect, the electric thrusters substitute for the lower efficiency chemical rocket engines on the launch vehicle. Electric type engines have too low a thrust to reach orbit by themselves, but by attaching them to a Skyhook, we can add orbital energy gradually, and then give that to the payload in a short time. The Skyhook becomes a very efficient battery for storing orbital energy, about 25 times the energy/mass of Lithium batteries on Earth.

Power Rquirements - For each kg of payload we are placing in orbit, we are changing the velocity from 5074 to 7474 m/s. This requires adding 15 MJ of enegy. Since there are 31.5 million seconds in a year, then that means for each kg/year of payload, we need 0.477 watts of delivered orbital enegy. In Earth orbit we are not in sunlight 100% of the time, so solar panels would need to be larger to average this power level, and electric thrusters are not 100% efficient, either. Using reasonable numbers of 60% sunlight time and 65% thruster efficiency, we get our solar panels need a peak output of 1.22 Watts/kg/year. If we want to deliver 1000 tons/year, then the power supply needs to be 1.22 MW. Nuclear power is excluded from consideration for political and safety reasons so close to Earth. Existing solar cells, allowing for 100% overhead, have a mass of 1.68 kg/m^2[3], and an efficiency of 29.5% vs a Solar constant of 1360 W/m^2.Therefore they produce 238 W/kg of output and their mass is 5.1 tons.

Thruster Type - For low Earth orbit, there are three types available with near-term technology: Ion, Plasma, or Electrodynamic. Electrodynamic uses less "fuel", but is not as well developed. Ion is well developed, but does not scale up to the high power levels as well. We will assume Plasma thrusters, but development of Electrodynamic should be pursued, and all three types considered as candidates. Current plasma thrusters in development are designed for 200 kW continuous power, so 6 units plus some number of spares would be needed for the Skyhook design.The estimated mass of the thrusters is 3 tons.

Fuel Requirement - With 1000 tons/year of cargo delivery to which 2.4 km/s of velocity is added, we need to expel 48 tons/year of thruster propellant at 50 km/s to maintain the Skyhook orbit. This can either come along with the cargo, serving as 4.8% overhead, or if materials are being extracted from nearby asteroids can come from them. The latter is preferred since it's more efficent for the launch vehicle.

Design Issues

Space Environment

Radiation Environment

Plasma Environment - The ionosphere can cause charge build up.

[Text still to be integrated]

multiple smaller orbiting systems, known as **Rotovators** can perform most of the same task, and use current materials.Each rotovator has a smaller task, so the required material strength is lower. We also take advantage of orbital mechanics to transit between them, which requires no materials at all. Various space elevator concepts have been theorized, and small-scale experiments flown in space. Significant R&D is needed to bring this technology to a ready state. When traffic volume is not large, and much of it is restricted to low orbits, the savings from a space elevator are not large enough to justify their construction. We therefore place it in this sub-phase, where reaching high orbits gives them greater advantage. Elevator research can be combined with a variable gravity research facility, as both can use rotating structures. An eventual skyhook network can provide fast velocity changes for people and cargo around gravity wells, while electric tugs can perform slower transfers between them, and to new locations beyond the existing network.

Skyhooks

A space elevator system in the form of a rotating Skyhook would allow using highly efficient electric thrusters in place of low performance chemical rockets for much of the transport job in gravity wells or between orbits. The first one could be built in Low Earth Orbit, and then others in higher orbits and around other bodies. The Earth's gravity well is too deep to fully span with current materials, so the low orbit Skyhook is not a full ground-to-orbit elevator. Still, reducing the work for a launch vehicle by 30-50% brings dramatic cost reductions.For smaller bodies such as the Moon or Mars, a Skyhook could span the whole gravity well.

As a large transport infrastructure project, similar to a bridge or airport on Earth, the Skyhook is built when traffic demands it and not before, and then expanded incrementally. The materials for the Skyhooks, such as carbon fiber, may come from orbital mining and processing. In that case their construction would not require large amounts of mass to be launched from Earth. Even if all the mass has to be brought from Earth, the potential for improved payload justifies at least more analysis to see if it is feasible.

1. Gunn, Dr. B.M. **Lunar Basalts and Anorthosites** (http://www.geokem.com/lunar.html)
2. National Academies Press **Orbital Debris: A Technical Assessment, 1995** (http://www.nap.edu/catalog.php?record_id=4765)
3. Emcore Space Solar Cell Products (http://emcore.com/solar_photovoltaics/space_solar_cells)

Section 4.8: Phase 4 - Orbital Development (continued)

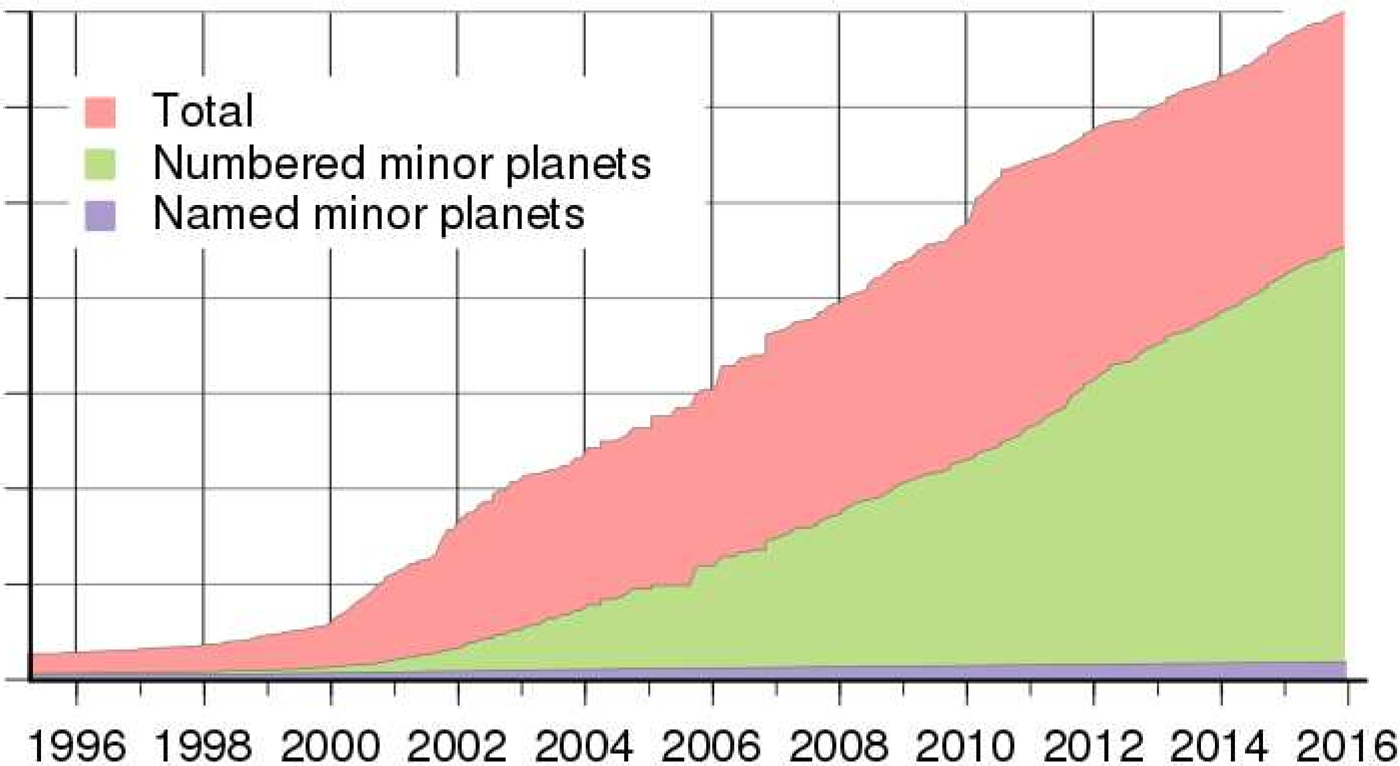
The remaining parts of Phase 4 (4D, E and F) are too undeveloped at present to devote full sections to them. For now we will gather our early ideas about these program phases here in one place, pending further concept exploration work.

The earlier parts of Phase 4 cover developing the Low Orbit (Phase 4A) and High Orbit (Phase 4B) regions around the Earth, and the Inner Interplanetary region (Phase 4C) around the Sun to a distance of 1.8 AU. They were described previously in sections 4.5 to 4.7 of this book. Those three regions are the closest and easiest to reach from Earth, so we expect them to start development first. The remaining three orbital regions cover successively farther regions from the Sun: Phase 4D - Main Belt and Trojan, Phase 4E - Outer Interplanetary, and Phase 4F - Scattered, Hills, and Oort. These are farther in the future, so our concepts are less developed for them.

Phase 4D - Main Belt and Trojan Development

The first known asteroid, **1 Ceres**, happened to be discovered on the first day of the 19th Century - 1 January 1801. Through that century 462 more were discovered, and by 1951 the count had reached 2,158. Since then, larger telescopes, electronic sensors, and automated analysis have greatly increased the known population. It reached 28,000 by 1995, 280,000 by 2005, 750,000 by 2017, and is still rapidly increasing (Figure 4.n-1). Their locations were originally concentrated in what we now call the **Main Belt** between Mars and Jupiter. Asteroids are now known to exist all over the Solar System, from inside the orbit of Mercury to far beyond Neptune.By count, the largest number are still located in the Main Belt, but this may be observational bias.Ones that are farther from the Sun are dimmer, so we tend to only find the larger ones. Ones that are closer to the Sun are hard to see due to interference by the Sun itself, and the fact we are looking at their unlit side.

Although asteroids occur everywhere in the Solar System, for program purposes we divide them into four regions by distance from the Sun, with a separate phase for each. This is due to variations in solar flux, temperature, and other environment parameters, and differences in average composition. These features will drive different designs for each region. Development of the Main Belt and Trojan region is an



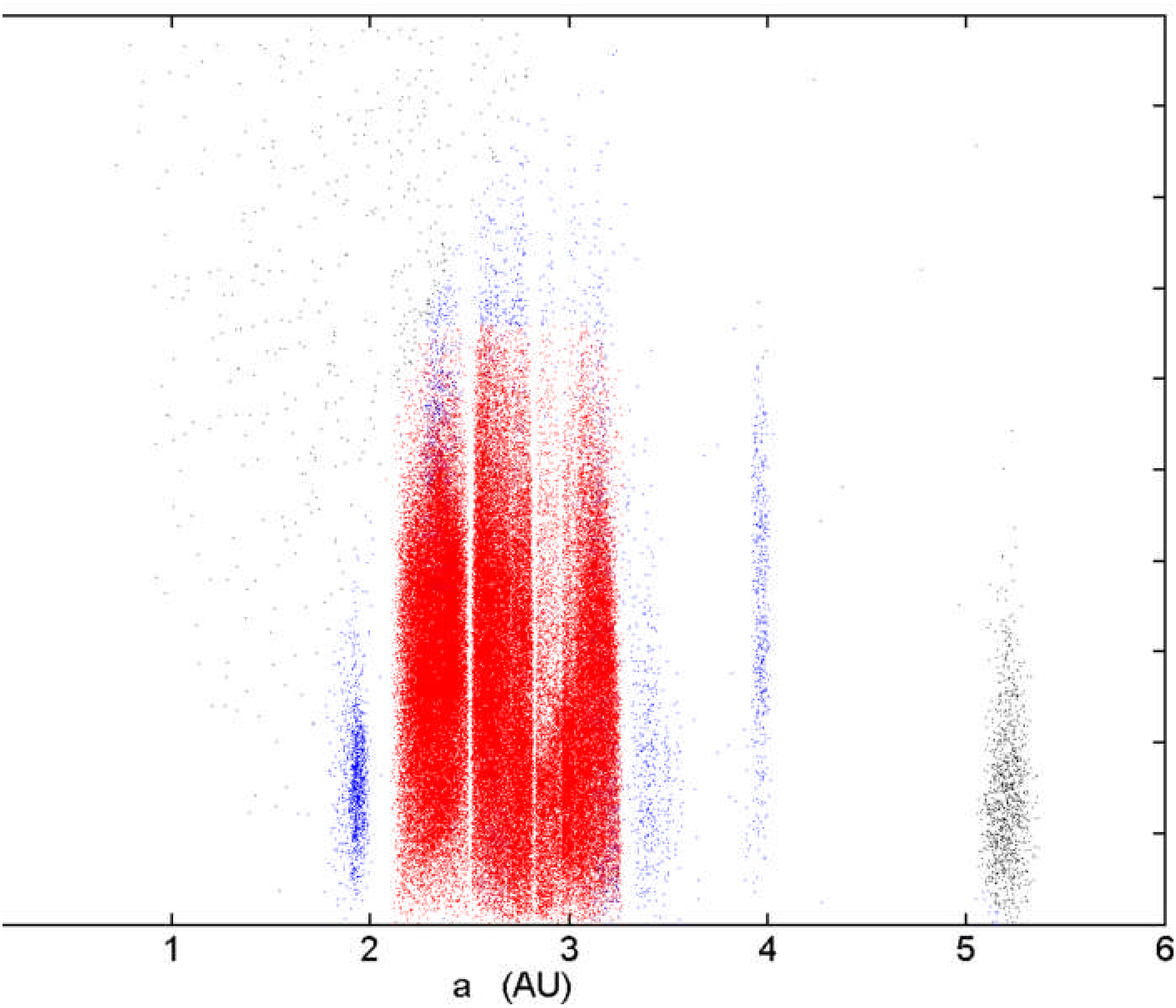
Total number of discovered, numbered, and named minor planets (asteroids) since 1995.

extension of work in the Inner Interplanetary region of Phase 4C, which starts earlier because it is closer to Earth. Both regions have objects with a range of orbit eccentricities (Figure 4.n-2). So each object varies in distances from the Sun and orbits as a whole overlap, making the region boundaries fuzzy. We set an inner limit for this region just beyond Mars at a semi-major axis of 1.8 AU, where the density of Main Belt asteroids significantly increases, and the outer limit at 5.4 AU, where the density of the Jupiter Trojan group falls off. This is an arbitrary choice, but it includes a very large number of smaller bodies in similar environments. Therefore we can develop a shared set of designs across the region.

Region Features

The region includes the Main Belt asteroids, with a core region between 2.1 and 3.3 AU where their density is highest. It also includes the **Hilda Group** which are in 3:2 resonance with Jupiter. Their orbits are between 3.7 and 4.2 AU from the Sun. The final major group are the **Jupiter Trojans** which occupy the Lagrange regions ahead of and behind Jupiter. They share the same average distance from the Sun as Jupiter, in the range of 5.2 ± 0.15 AU. There is a relatively small percentage of the total region population that doesn't fall into any of these major groups. The region does not include Jupiter itself and orbits within 20 million km of the planet (see Phase 5D, below).

Historically, asteroids and comets were regarded as separate classes of objects. We now know some objects are actually former comets which have lost most of their volatiles, and now look like asteroids. Some objects identified as asteroids are still releasing vapor, notably including the largest one, Ceres. It is therefore reasonable to consider all small bodies in the region as a single class, with a range of compositions and solar distances. We will use the name "asteroids" for all of them, since they are by far the largest in number. Objects traditionally called short period comets, with semi-major axes between 1.8



· Plot of eccentricity (e) vs semi-major axis (a) for numbered asteroids closer than 6 AU.

and 5.4 are included under our asteroid heading, but make up only 0.1% of the total population.

Asteroid sizes range from 945 km in diameter for 1 Ceres, which is now counted as a dwarf planet, down to **Interplanetary Dust** of sub-millimeter scale. The dust component is short-lived and does not account for much of the mass in the region. About one million tons/second of solar wind particles flow through the region at high velocity, but the flow is very diffuse, on the order of 1 nanogram/km^2/s. Total mass in the region is about 3 x 10^{18} tons, which is equivalent to 16 million years of Earth's total current mining output. All of it is available to determined mining efforts, because the low gravity on the asteroids creates low subsurface rock pressures. About half the total mass is in the four largest objects: 1 Ceres, 4 Vesta, 2 Pallas, and 10 Hygeia.

Asteroid compositions vary considerably due to differences in their formation and history (**DeMeo, 2015**). Between **spectroscopic observations**, and **examination of meteorites**, many of which are fallen pieces of asteroids, we can identify a number of composition groups. However, only a few asteroids have been visited by spacecraft, so detailed verification of their compositions is yet to be done in most cases.

Velocity to reach orbit from the largest body, Ceres, is only 270 meters/second, or 860 times less kinetic energy than from Earth. So materials from these asteroids are easy to export once you are near them. The main energy cost is changing orbit around the Sun to reach them. Solar power is available 100% of the time in the region, except shadowed areas around and on the asteroids. Intensity varies from 31 to 3.4% of that near Earth. Ambient temperature varies from 244 to 217K (-29 to -56C) for black objects, and less for lighter colored ones. Travel time from Earth is typically years using least energy trajectories, with high to lethal radiation levels for unprotected people. Communication time from Earth varies from 13 to 120 minutes round-trip, including a relay when needed to avoid a direct path through the Sun.

Development Projects

This region is nearly devoid of spacecraft at present, so most uses are in the future. Abundant raw materials of diverse composition, and adequate amounts of energy when concentrated, will enable mining as an early activity. Materials would be shipped to previous regions at first, which are more developed and have higher solar intensity for further processing. When it makes sense to do so, seed factories can help bootstrap a full range of local industry, and eventually large scale habitation. There is enough total raw materials and energy in this region to support a full civilization.

The largest object in the region is the dwarf planet **1 Ceres**. Equatorial orbit velocity is 359 m/s, and equatorial rotation velocity is 94 m/s. So to reach orbit requires 265 m/s net. This velocity can be reached by a mild steel centrifuge, and easily reached with any advanced material. Therefore bulk material launch from any other Main Belt asteroid, all of which are smaller than Ceres, does not require any rocket propulsion. A 1-g Skyhook would be 7 km in radius for Ceres, and allow crew and equipment to be landed and take off from at low acceleration, and a cost of 0.5% of net mass flow in reaction mass to maintain orbit. So surface access for any asteroid should not be difficult. For the smaller bodies, the operation is closer to docking in zero gravity than landing from orbit.

For smaller asteroids staying on the surface will be more of a problem than getting on and off. For example, the 35th largest asteroid by diameter is **9 Metis**, which has an equatorial radius of 170 km and a mass of 1.47 x 10^{19} kg. This gives a surface gravity of 0.034 m/s^2 (0.34% of Earth). The rotation period is 5.08 hours, which give a rotation velocity of 58.4 m/s and a centrifugal acceleration at the equator of 0.020 m/s^2. So the net effective gravity is only 0.014 m/s (0.14% of Earth gravity). Indeed, the orbital velocity is 76.0 m/s, so it only takes 17.6 m/s (39 mph) added velocity to reach orbit. Therefore humans or low speed machinery can toss things into orbit, and a firm anchoring method will be needed to not have equipment move accidentally.

Production

The inner parts of the region have enough sunlight for solar panels to produce power directly. In the outer portions, solar panels will benefit from reflectors to increase the light intensity. Concentrating reflectors can produce high temperatures at all distances, either for industrial processes or warming habitats. Increasing amounts of reflectors are needed as you get farther from the Sun, but they are inherently low mass in a zero gravity environment with no weather. Note that the total amount of solar energy available in this region is no larger than for the Inner Interplanetary region, and equal to the total output of the Sun, which is 3.83 x 10^{26} W. It is the same total solar flux, only more spread out as it increases in distance. The difference is the Main Belt & Trojan region has more raw materials available than the inner region.

Asteroids are generally covered in a mixture of rocks and dust of varying sizes. This is the result of repeated impacts over their life and gravitational attraction of loose orbital material. In fact, some asteroids are so low in density that they must be "gravel piles", with no solid central body. The loose material makes surface mining easy, but at the same time most asteroids are small. The rocks and dust are easily disturbed and can become a hazard to mining and production operations. So attention has to be given how to carefully remove materials without too much disturbance. They are then moved elsewhere by a tug, or to a nearby processing plant out of range of any dust clouds created. For larger operations, an inflatable or assembled shell can surround the whole asteroid, keeping dust contained. Processing equipment can then be attached to the outside of the shell, and materials delivered continuously until the asteroid is consumed. Because dust and debris is contained, more vigorous mining methods can be used. Mining and processing methods should have been developed earlier for the Inner Interplanetary region due to similarity in asteroid sizes and types. The one difference is the larger size of some bodies in the Main Belt and Trojan region, making their gravity significant enough to matter in design.

Habitation

Habitats for this region can start with unmodified designs from the previous regions, except with the addition of reflectors for increased power and keeping warm. Early units can be delivered whole from inner regions and moved gradually into this region over time. With continued access to nearby asteroids for supplies, there is no need to deliver them all at once. Once in place at a good location, such as orbiting Ceres, an early habitat can grow by making and assembling structural parts for larger habitats, then a series of shells of increasing size. Ceres is in the middle of the densest region of the Main Belt, so supply trips to nearby asteroids with different composition will be relatively easy. This makes it a good candidate for starting large-scale development of the region.

Transport

The same transport methods can be used in this region as for the Inner Interplanetary region. The main difference is adding reflectors to solar panels, or larger reflectors to thermal power units, to make up for the lower solar intensity. Centrifugal transport hubs are somewhat more efficient for injecting bulk cargo to transfer orbits, because they do impulse transfers rather than spiral orbits. If a large asteroid absorbs the reaction force, they also don't need propellant to send cargo on their way.

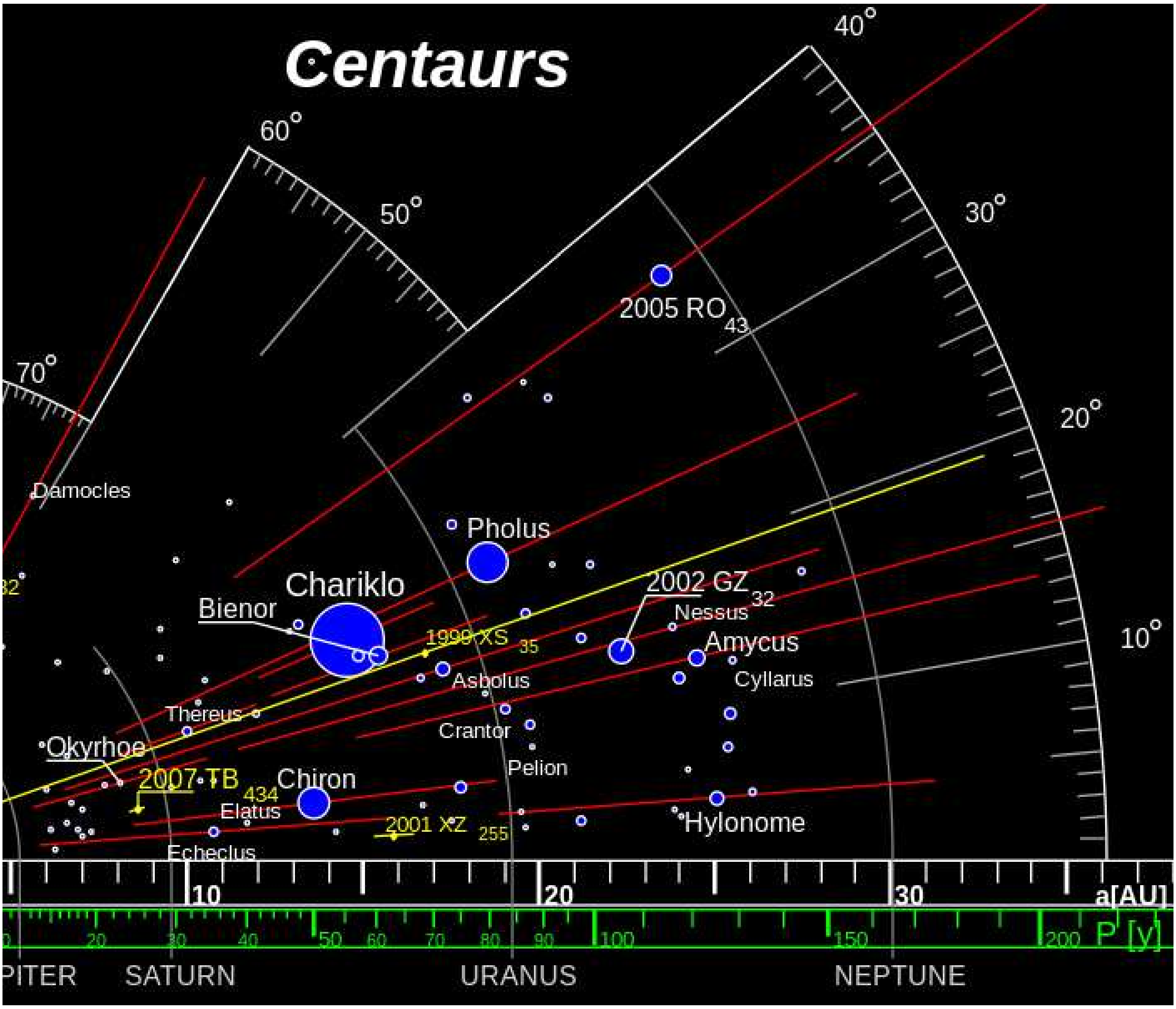
Services

The first service functions in the region will be communications, scientific exploration, and prospecting, to locate and define available resources in detail. Other services are to be determined later

Phase 4E - Outer Interplanetary Development

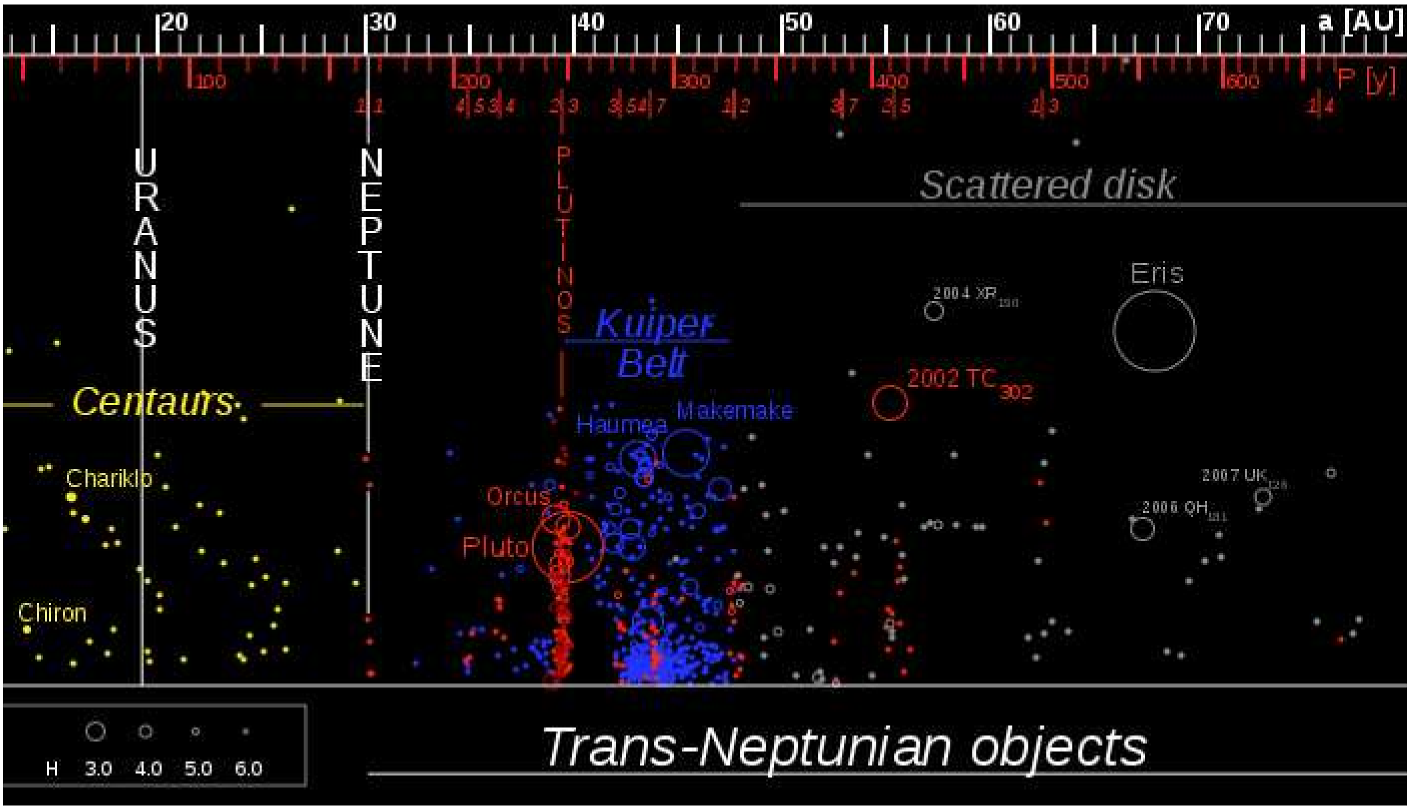
The Outer Interplanetary region is the third such region to begin development, after the Inner Interplanetary (Phase 4C) and Main Belt and Trojan (Phase 4D) ones. It is the next in physical distance, covering orbits with semi-major axes from 5.4 to 50 AU. It excludes the major planets Saturn, Uranus, and Neptune, their moons, and an orbital region around each, which are assigned to Phase 5E. Only a few spacecraft have reached this region by 2017, and most were directed at the major planets and Pluto, so it is largely unexplored. Most of our information to date comes from astronomical observations on or near Earth.

Region Features



stribution of Centaur asteroids by distance a (AU), period P (years), inclination i (degrees), and size (circles). The distance s are marked by arcs.

The first object found in this region was the dwarf planet **134340 Pluto**, in 1930. As of late 2017 the known population has grown to about 275 in the **Centaur** group (Figure 4.n-3), and nearly 1800 in the



stribution of Outer Solar System objects to 80 AU by named groups (color), distance a (AU), period P (years), inclination i
ize (circles).Distances of the Gas Giants are marked by white lines, and resonant orbital periods with Neptune in red
.

Kuiper Belt group (Figure 4.n-4) beyond Neptune. The Centaurs have orbits between or cross those of the four Gas Giants, including Jupiter. This tends to make their orbits unstable and short-lived. The Trans-Neptune group as a whole spends most or all of their time farther than Neptune, so their orbits are more stable. The large number of objects in the inner part of the Trans-Neptune group, from 30-50 AU, are referred to as the Kuiper Belt. There are also about 20 known **Trojan** objects whose obits are tied to the outer Gas Giants, mostly Neptune, and about 240 short- and long-period comets in the region.

For our purposes we group asteroids and comets in the region together as forming a continuous range objects with varying orbits and compositions. Comets are distinguished by sometimes coming close enough to the Sun to actively lose gas and dust. Historically this made them easy to spot. But at other times they are as inactive as asteroids, which keep more consistent distances. Current telescopes have a hard time finding inactive objects in the region which are smaller than 10 km in diameter, so our count is incomplete and continues to grow.

About six of the known objects in the region (Pluto, Makemake, Haumea, Orcus, Quaoar, and Varda) are large enough to be considered dwarf planets, and about 675 are estimated to be larger than 100 km in diameter. Total mass in the region is estimated at 240-600 x 10^{18} tons (4-10% of Earth), which is a very large amount of total available material. Except for the deeper parts of the largest bodies, most of this is theoretically accessible. Due to generally low gravity and density, the subsurface pressures are not too high for mining operations, and the interior temperatures are probably not too high to be a problem. The Centaurs are likely to be of mixed composition. Since their orbits are unstable, they originally came from elsewhere, where the conditions of their formation were different. Water ice and carbon compounds have been detected on several of them.

The entire region beyond Jupiter is outside the **Frost Line**, the distance in the original Solar Nebula where water ice could condense. Therefore water is common in the region, and other ices, like methane, ammonia, and nitrogen, are present in the outer portions, where the local temperatures were cold enough for them to also condense. Since the Solar System's formation, the opaque Solar Nebula has dispersed, and the Sun has gradually brightened, increasing temperature at a given distance. So surface materials which were originally stable have since evaporated. They can survive to the present deeper within objects. Changes in orbit since their formation will also have affected what remains in these objects. The larger bodies have undergone impact heating during formation, and radioactive heating afterwards. This causes them to separate into layers by density. Nominally this would be metallic and rocky material towards the center, and icy material towards the surface. Smaller impacts and exposure to solar ultraviolet and other radiation may have modified the surface layers. Since very few of these objects have been explored close-up, we can only speak in generalities at present. Much more exploration and prospecting will be needed before we can start to use the materials from this region, and begin local development.

Escape velocity from Pluto is 1.2 km/s, and less for smaller objects. This is well within the reach of mechanical transport. The minimum velocity to reach the outer parts of the region from Earth is near Solar System escape, or 12 km/s. Such orbits will take over 60 years, so faster transport using more energy is desirable. The dominant energy cost in using the region is then first reaching it. Gravity assists and advanced propulsion will be needed to access the region in reasonably short times, or a lot of patience.

Available total solar power is the same as for the previous two regions, being the total output of the Sun. The intensity per area, however, is low, from 3.4% to 0.04% of Earth orbit values. This would require large reflectors to increase intensity, or using nuclear or other power sources instead. Ambient temperatures are very cold, from 217 to 70K for black objects, and lower for lighter colored ones. Travel time from Earth will typically be many years. Unprotected radiation levels are high to occasionally lethal for people, and damaging over long periods for equipment. Round-trip communications time is 1.2 to 13.6 hours on a direct path, and slightly higher if a relay is needed to avoid the Sun.

Development Projects

This region is likely too far to do much beyond scientific exploration with present technology. When civilization has expanded through the previous regions, and better technology is available, the first use is likely to be mining the large sources of raw materials. They would be brought back to inner regions, where there is more energy density to process them and use them for other activities. The combination of low temperature and sorting by density makes the various ices the most accessible early resource. Use of this region is far enough in the future that technology is likely to improve dramatically in unexpected directions. So any concepts we present for this region should be considered very preliminary and likely to change.

Production

We don't expect a lot of production besides mining in this region until technology significantly improves. Ices like water and nitrogen are very useful to people, and available in large amounts in the region. So mining and transport to the inner regions is a possibility once there is enough demand. Transport would be slow, taking many years, so there is an incentive to set up a "pipeline" of cargo in transit, with vehicles at each end to set it on course and collect it at the end. The cargo can travel unattended in between, saving on vehicle time. Once the pipeline is filled, then cargoes arrive on a regular schedule. If fusion is well developed, a fusion-based economy may develop, with full production and habitation. We don't see a strong reason to live this far out, rather than staying in the warmer and brighter inner regions, but such reasons may develop.

A challenge for the Kuiper Belt and farther regions is supplying enough solar energy to operate. Civilization on Earth consumes about 2.7 kW/person, and we would expect a higher number for space locations, both due to higher standard of living, and the need to do artificially things handled by natural processes on Earth. Let us assume 20 kW/person is needed, system mass is double that for the ISS, or 150 tons/person, and half is devoted to solar collection. If magnesium-aluminum reflectors 1 micron thick are used to concentrate sunlight, they will have a mass of 2.4 tons/km^2. So we are allowed a maximum of 31.25 km^2 of reflectors/person. For a net power of 20kW at 1/3 efficiency, we need 60 kW of sunlight. At Earth, solar flux is 1.361 kW/m^2, so we need 44 m^2. Since we are allowed 711,500 times this area, and solar flux falls as the inverse-square of distance, we can provide sufficient solar energy out to 843.5 AU, a surprisingly large distance. Beyond this, operations would limited to low power situations, or require other sources, like nuclear or beamed energy.

Habitation

[TBD]

Transport

Due to weak sunlight in this region, we expect that nuclear powered propulsion, and gravity assists from the larger bodies, would be major ways to get around. If nuclear fusion has not been sufficiently developed, fission would be the only available nuclear source. There is a finite known supply of suitable radioactive elements on Earth and the Moon. To supplement them, artificial radioactives can be produced near the Sun, where abundant energy can power accelerators to convert non-radioactive starting materials. If nuclear fusion is well developed, there is abundant hydrogen in the region from which fusion fuels can be extracted. As distance increases from the Sun, orbit velocities, and so the required orbit velocity changes, decrease as the square root of distance. Solar flux decreases faster, as the inverse square of distance. So solar sails become will become less effective as a transport method than for closer regions.

Services

[TBD]

Concept Details

We note a few features about **136108 Haumea**, a large object in the outer part of the region. Haumea is massive enough to be in hydrostatic equilibrium, and therefore is classed as a dwarf planet. However, the short rotation period (3.9155 hours) means it is not round, but rather ellipsoidal, with a long axis about twice that of the short axis. Circular orbit speed at the long ends is ~527 m/s, while the tips themselves rotate at ~428 m/s. So only ~99 m/s velocity change is needed to land or take off from it, one of the lowest numbers for a large Solar System object. If Haumea retains any sort of atmosphere, it would tend to be in a wedding-band shaped ring around the short axis. Gravity would also vary significantly from the long ends to the short axis.

Phase 4F - Scattered, Hills, and Oort Development

The vast space beyond the Kuiper Belt is the fourth and last interplanetary region to begin development. It includes orbits with semi-major axes from 50 AU to the limits of the Sun's gravitational dominance, which we set at 100,000 AU. Although it covers a huge range of distances, it is a small range in energy when measured from Earth, covering the last 2% relative to reaching solar escape. Only four spacecraft have entered this region by 2017, after completing their primary missions closer in, with a fifth to enter it in a few years.So nearly all of our information comes from observations on and near Earth.

Region Features

Long-Period and **Near-Parabolic Comets**, whose orbits are large enough to be counted in this region, have been visible since ancient times. Determining that their orbits were in fact so large had to await the development of orbital mechanics and better telescopes. Active comets are easily seen when close to the Sun. They emit large amounts of gas and dust when heated, creating a coma and tail which can extend millions of kilometers.When far from the Sun, they are cold, dark, and inert, and therefore much harder to find. So the first object in this region that wasn't an active comet, **(48639) 1995 TL8**, was not discovered until 1995, and that one because it is relatively large - about 350 km in diameter for the primary and 160 km for its satellite.

The known population of objects in the region (as of late 2017) includes 100 long-period and 420 near-parabolic comets, and 440 **Scattered Disk Objects**, whose orbits lie entirely beyond Neptune and are therefore relatively stable. Their name comes from being scattered by the major planets out of closer orbits in the Solar Nebula where they formed. Four known objects have maximum distances greater than 2000 AU. This places them in the Hills Cloud, whose orbits range from 2000 to 10,000 AU in aphelion. This is presumed to be a large reservoir of objects scattered farther away, and limited by closely passing stars in the cluster where the Sun formed. Beyond this is the Oort Cloud, which extends to 100,000 AU in semi-major axis. We have indirect evidence for a large population in the outermost areas, based on the orbits of known near-parabolic comets. Oort Cloud objects are far enough from the Sun to be affected by galactic tides and passing stars and massive gas clouds. These forces sometimes send them close to the Sun, where we see them as active comets.

Our ability to detect all these distant objects is currently limited to the larger ones which are presently within about 80 AU of the Sun. So our discoveries over the last 20 years come from objects in the region whose closest orbital distance (perihelion) is less than 80 AU, and which happened to be near that minimum distance. Since orbit velocities are lower at greater distances, the ones with highly elliptical orbits spend most of their time too far to see. The ones with more circular orbits which stay more than 80 AU from the Sun can't be found at all today, and neither can most of the ones less than 15 km in diameter. We therefore expect to find many timesmore objects in the region as our instruments impove.

Total mass in the region is poorly known at present, but is estimated to be 4-80 times that of Earth, which is a vast reservoir of materials. This total includes a suspected, but as yet undiscovered, 9th planet with a possible mass in the range of Neptune's (~15 x Earth's mass). Since comets are from this region, and their evaporating gas and dust is easy to observe, we have a reasonable idea of compositions in the region, even though we can't directly observe most of it. It is most likely a mixture of water, other ices, complex carbon compounds, and some heavier mineral grains. Roughly 4% of the population would have originated from inner parts of the Solar Nebula, and therefore be more rocky or metallic than volatile compounds. Solar energy is quite weak in the region, below 0.04% of that near Earth, and ambient temperatures are below 70K down to near 2.7K. Travel time with current propulsion technology is many years to centuries. Round-trip communications time ranges from 14 hours to 3 years. These plus the required orbital energy to get there make reaching and working in the region difficult, despite the large amounts of material likely to be there.

Development Projects

We don't have enough information about objects in this region to make detailed plans, and they are too far away to access with current technology. So anything beyond science and exploration are deferred to a future time when increased needs and better technology exist. When that time comes, though, there is a very large reserve of materials from the region that can be put to use.

Production

We showed under Production for Phase 4E that enough solar energy is available even to 1000 AU from the Sun to sustain production and habitation. Beyond that, nuclear or beamed energy sources would likely be needed. Production beyond materials extraction must remain speculative at present.

Habitation

[TBD]

Transport

To get transport times to the region to reasonable levels, very high energy propulsion would be needed, such as nuclear fusion. Since the light elements needed for fusion are common in these outer regions, this could be self-fueling once set up. Unfortunately, fusion is not yet a viable technology, so transport that uses it remains speculative at present.

Services

[TBD]

Section 4.9: Phase 5A - Lunar Development

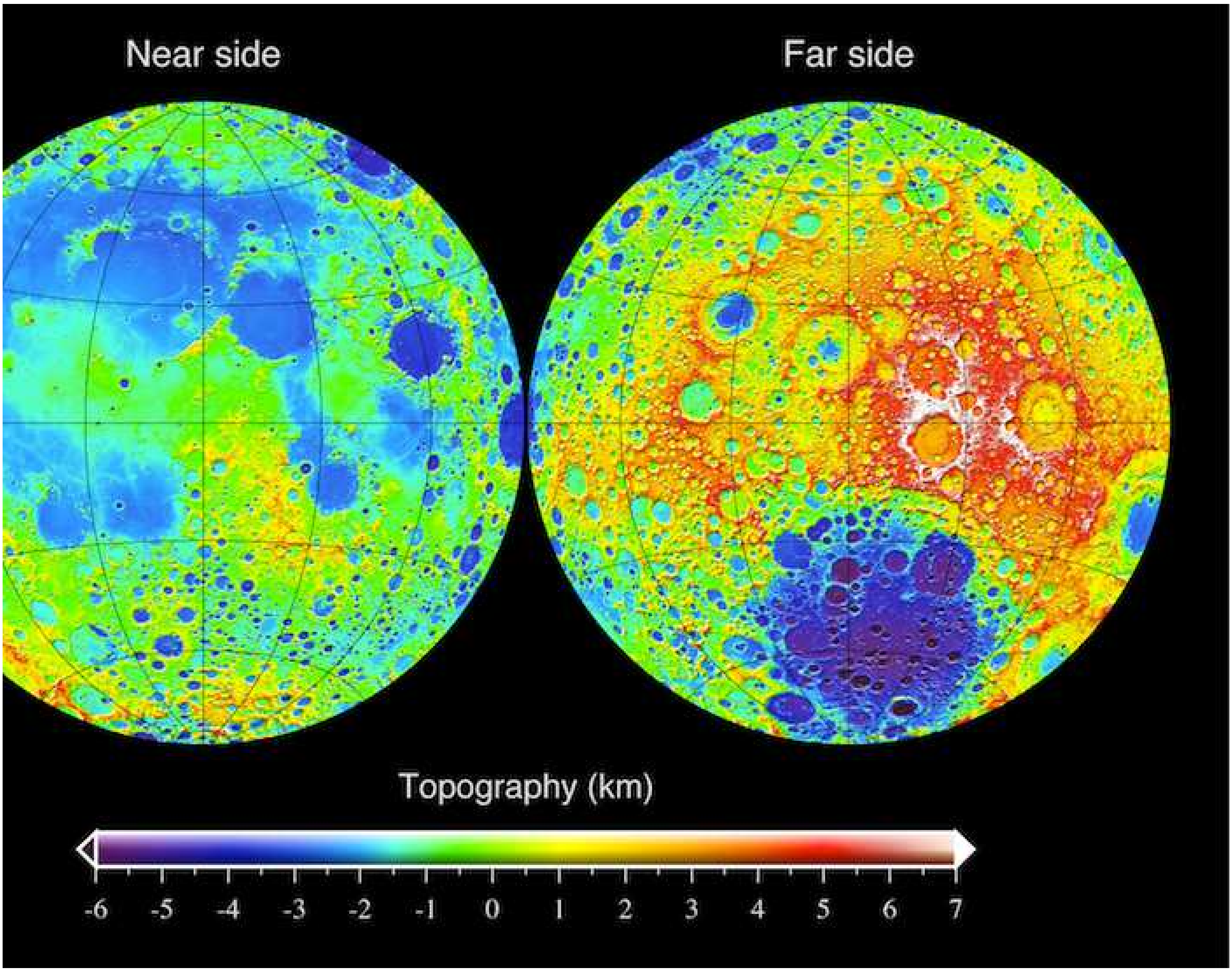
The **Moon** is obvious to anyone who has looked up at the sky. The same reasons for expanding civilization apply to it as to other locations. They include access to new energy and raw material resources for upgrading civilization, and to meet the other objectives listed at the start of **Section 4.1**.

However, we do not start developing the Moon right away for several reasons. First, we must travel through the Low and High Orbit regions to get there, then navigate the Moon's gravity field to reach closer orbits or the surface. This requires more energy and new transport equipment than the orbital regions. Second, the surface has different conditions than the orbital region around it, or the orbital regions around the Earth. So surface equipment also needs new designs. Third, developing the Moon is made easier with support from the previous orbital locations. So we delay the start of Lunar development until after Phases 4A and 4B have been started, and needed (R&D) for this phase is completed in **Phase 0M**. The Moon's surface is roughly the size of Africa and Australia combined, so it is much too large to be developed all at once. Once started, Phase 5A will continue in parallel with previous phases, with progressive upgrades and expansions over time, both in orbit and on the surface.

Our concept exploration for developing the Moon begins with describing the characteristics of the region and an industry survey to identify possible future activities. Motivations, economics, technology, and prior developments elsewhere will drive which of these activities can get started and when. Combining this information we can identify a development approach and specific Lunar projects, place them in approximate time order, and link them to other phases and projects. In the course of doing this, we identify needed R&D for Phase 0M, which is recorded in that section.

Lunar Region Features

The Lunar region includes the Moon itself, and orbits with average distances (semi-major axis) within 35,000 km of the Moon's center. At this distance the Moon's gravity exceeds the Earth's by 50% and is the dominant local force. Lunar orbits in general are unstable due to the influence of the Earth and Sun. The Moon also has mass concentrations from past impacts that create an uneven gravity field. Particular orbits, where these effects are minimized, can be stable for long periods. Otherwise propulsion must be used to correct them, or the orbit allowed to evolve over time. Since the Moon orbits the Earth, the Lunar region is embedded in the larger High Orbit region around our planet.



Lunar topography referenced to average radius (a 1737.4 km sphere).

As noted above, horizontal surface area of the Moon is quite large, 37.93 million km^2 or about one quarter of the Earth's land area. This does not include the sloped exposed area from large-scale topography (Figure 4.12-1) and numerous craters and other small-scale features. See **USGS Topographic Map 3316 of the Moon, 2015** for a more detailed version.

The Moon orbits the center of the Earth-Moon system every 27.3 days with respect to the stars (inertial frame), and 29.5 days with respect to the Sun, which sets the length of the Lunar day. The Moon makes one rotation per orbit period, and therefore keeps approximately the same side facing Earth. It is not exact because the Moon's orbit is not circular, it has a slight residual pendulum motion, and the Earth's diameter is 1/30 of the Moon's distance. That size allows varying views of the Moon, depending on where you are. Orbit durations around the Moon vary from 108 minutes close to the surface, to 6.8 days at the upper edge of the region. The main environment and resource features of the region are as follows:

Environment Parameters

Temperature - The Moon has the same average distance as the Earth from the Sun, so the incoming **Solar Flux** of 1361 W/m^2 is the same. That sunlight is partly blocked by the Moon's shadow in lower lunar orbits, and blocked 50% or more on the surface on a monthly cycle. There is no significant atmosphere to moderate temperature changes on the surface. Objects exposed to space and direct sunlight can vary from 100-400K at the equator, and as low as 30K in shadowed polar craters. The surface is covered by particles separated by vacuum, and is an excellent insulator. Soil temperatures more than 30 cm deep are not expected to vary more than a few degrees from average, despite two week intervals of daylight and night. The average soil temperature is about 240K at the equator (**Vasavada et al. 2012**), and colder with increasing latitude and lower sun angles. Heat flow from the hotter interior has been measured at 0.008-0.03 W/m^2, depending on location. The concentration of radioactive elements in the crust and good insulation properties leads to a somewhat uncertain thermal gradient of 10-50 K/km with depth.

Temperatures in orbit will mainly depend on the percentage of time in sunlight vs in the Moon's shadow, and how much light is reflected or emitted from the Moon at a given distance. The Earth appears about 2 angular degrees in size from the Lunar region, and thus fills less than 0.01% of the sky, and makes a small thermal contribution. The remainder of the sky is the cosmic background at near absolute zero (2.7K), which serves as an effective heat sink. Since orbits can be elliptical and change over time, the resulting thermal environment can vary significantly. The albedo, emissivity, and thermal properties of an orbiting object will determine how much heat is gained and lost from environment conditions, and what equilibrium temperatures they will reach.

Atmosphere and Water - As noted under temperature, the **Moon's Atmosphere** barely exists. Its total mass is less than 10 tons, which is the Earth's atmosphere over a single square meter. If pressurized environments are needed for people or equipment, the gases need to be produced locally, or imported. About 50 kg/s of solar wind particles flow through the region at high velocity. Since this is spread over an area of 3.85 billion km^2 it would be difficult to collect in useful amounts. An estimated 600 million tons of water ice has been discovered in permanently shadowed craters near the poles. The remainder of the Moon's surface has 10-1000 ppm of chemically bound water in minerals. The dry and airless state of the Moon is due to the low escape velocity, coupled with extended periods of early heating from impacts, tides, and radioactive decay.

Ground Loads - Ground strength for surface structures and transport is adequate to excellent. The surface consists of broken rocks and dust of various sizes, which compacts a few cm but then can support heavy loads. Excavation, mining, or drilling would reach the limits of overlying rock strength at ~38 km. Depths below this require support structures. Much of the shallower depths consist of fractured material, and will also need support against movement and collapse.

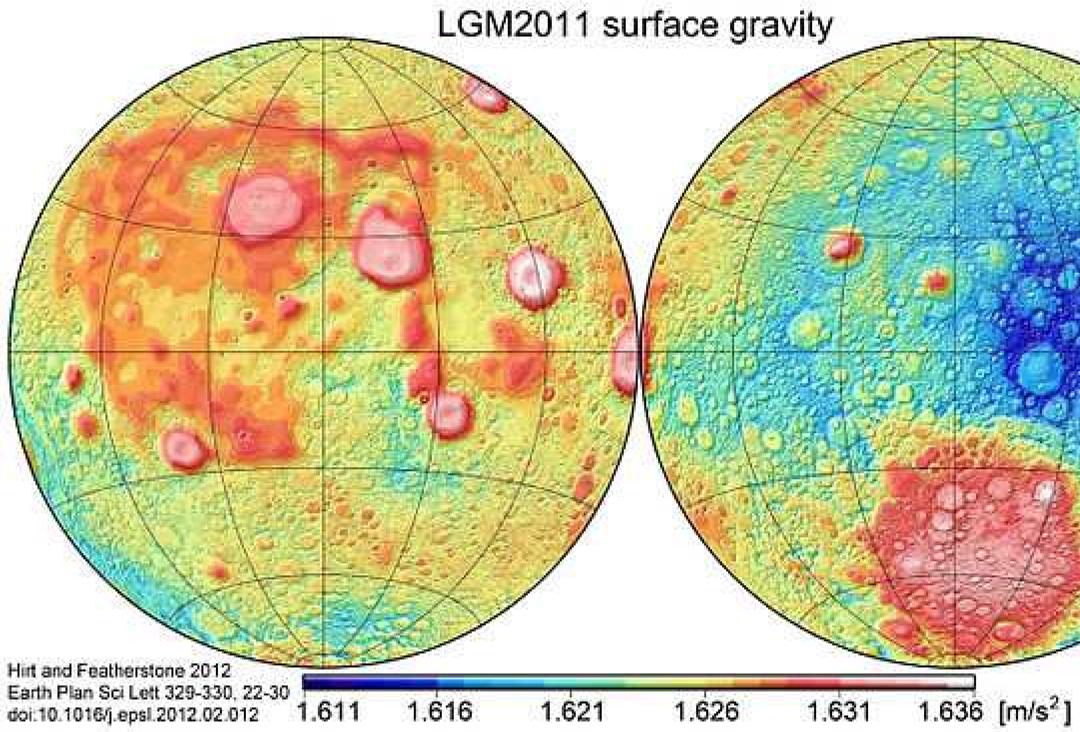


Figure 4.12-2 - Lunar surface gravity map.Near side on left, far side on right.

Gravity Level - Surface gravity averages 1.625 m/s^2, or 1/6 of Earth's, with a total variation of 0.0253 m/s^2 according to location (Figure 4.12-2). Free-fall conditions in orbit produce no effective acceleration. Structural support needed against gravity is therefore much lower than for Earth. The gravity levels required for long-term human health and plant and animal growth are not yet determined, but are likely to be more than natural surface or free-fall levels. Artificial gravity on both the surface and in orbit can be generated by rotation, otherwise stay times may be limited. Centrifugal forces from rotation will impose significant design loads on structures.

Radiation Level - Unprotected radiation levels on the surface and and in orbit are in the range of 0.1-1 mSv/day, which could reach lifetime astronaut exposure limits in a few years. A single powerful solar flare can deliver lethal amounts of radiation, although most are much weaker. Radiation also causes long term damage and electronic faults in equipment. The radiation mainly comes from the Sun and cosmic background. A meter or more of cover by lunar soil or building underground can provide safe radiation levels on the lunar surface.Bulk mass of various kinds can provide shielding in orbit.

Communication and Travel Times - Round-trip (ping) communication times from Earth to the Lunar region varies from 2.2 to 2.94 seconds, depending on where in the region, and the Moon's distance in its orbit. This includes satellite relay time if you are communicating with areas behind the Moon that can't be seen directly from Earth. Travel time from Earth is nominally 3-4 days for people, by direct transfer orbit. Cargo delivery by electric tug is much more efficient, but also much slower. Slow transits without shielding or other mitigation would expose people to high levels of radiation while traversing the Van Allen belts.

Stay Time - Average stay times for people affect transport needs and the types of habitation required. Only nine trips with people have been made so far to Lunar locations, and those lasted only a few days each. These trips were made nearly 50 years ago. This data is too sparse and too old to establish an average for design purposes. Future projects will need to define stay times based on internal needs, until long-term habitation establishes an average.

Transport Energy - Reaching the Lunar region from the Earth's surface involves first getting to a high orbit that intercepts the Moon's location, then entering a stable orbit or landing on the surface. The first of these theoretically requires 62 MJ/kg, and entering orbit or landing requires up to 2.8 MJ/kg. These are ideal values for potential and kinetic energy. The actual expended energy depends on the details and efficiency of the transportation methods. The transport energies from Earth are high compared to the 10-20 MJ embodied energy of typical products or 15 MJ/kg of high energy chemical propellants. This favors local production where possible.

Escape velocity from the Lunar surface is 2380 m/s, or 21% of Earth's. Therefore escape energy is only 4.5% of Earth's. Low orbit velocity is 1680 m/s, so the difference to escape is an added 700 m/s. Orbit velocity at the upper edge of the region is 375 m/s, and escape an additional 155 m/s. These velocities and the associated energy are relatively low. Internal regional transport and reaching the High Orbit and Inner Interplanetary regions are then fairly easy from this region.

Available Resources

Energy Supply - The Sun provides a nearly constant flux of 1361 W/m^2 at 1 AU. The Earth's orbit is slightly elliptical, and the Moon's orbit brings it closer to and farther from the Sun than the Earth. So the available energy in the Lunar region varies somewhat from this value. In full sunlight this energy is sufficient to process about 2 tons/m^2/year of raw materials. Collecting the energy with solar panels or concentrating reflectors may only require 2 kg/m^2 of equipment. So the primary energy return can be very high. Total solar energy flow through the region is 5.24 million TW, which is vastly greater than current world energy consumption of about 20 TW. Localized areas on the Moon contain up to 10 parts per million of Thorium and Uranium, which may be useful for energy production.

Most of the Moon's surface is exposed to the Sun 50% or less of the time, depending on local topography. Higher latitudes get less sunlight per unit of surface area due to lower Sun angles. Orbits are exposed to sunlight 50% or more of the time, generally increasing with altitude. Dark periods are relatively short in orbit. Dark periods on the surface last half a Lunar day, or two weeks. Since direct sunlight isn't available for this time, alternate energy solutions are needed. Possible approaches include:

- **Ceasing Operations**- High power operations can simply be stopped when sunlight isn't available, and minimal support provided by conventional sources like batteriesOne of the following options can be combined with reduced nighttime operations, but not stopping them entirely

- **Thermal Storage**- This uses the vast amount of rocks and dust as the storage medium, and the natural vacuum as insulation. The material is heated during the day by sunlight, and the heat used to generate power during the Lunar night.

- **Nuclear Power** - The surface is already bathed in high levels of natural radiation, so nuclear sources are not as much an issue as on Earth. A variety of nuclear designs are possible, with different power levels, and fueled from Earth or local sources.

- **Beamed Power** - Lunar orbits receive sunlight when the surface does not. Energy can be directed to the surface using simple reflectors, microwaves, or lasers, with varying beam target size vs distance.

- **Transmission Lines** - East-West transmission lines can deliver power from areas in sunlight to those that are dark. In the worst case at the Equator the lines would have to be 2700 km long to provide continuous power. This is 1/4 of the Lunar circumference, with one line in each direction. These are very long distances, and would only make sense when large-scale development is present. Shadowed craters get no sunlight, but the crater rim may get adequate power levels. A shorter transmission line or reflector can then deliver that power to where it is needed.

Materials Supply - Lunar orbits are essentially devoid of raw materials, so projects located there must import what they need. The Moon's surface has a reasonably well understood, **Geology**. See also the Lunar and Planetary Institute's **Lunar Sourcebook** (1991) for more detailed information. That understanding is from a number of lander and orbital missions, some of which returned samples for analysis, and from Lunar meteorites which were thrown to Earth by impacts. Broadly, the surface composition is oxide minerals of silicon, iron, calcium, aluminum, and magnesium, in order of elemental concentration, with 3-4% other elements. The Moon's surface is not uniform in composition (**Lawrence et. al. 1998**), with higher concentrations of Iron, Potassium, Phosphorus, Rare Earth, and Thorium found mostly in the Oceanus Procellarum region. There are few volatile (low boiling point) compounds left on the Moon. It formed in a molten state, then suffered many high energy impacts, and tidal and radioactive heating which kept it molten for extended periods. The Moon is too small to keep an atmosphere, so the volatile compounds mostly escaped. While molten, the denser materials sank to the interior, and lighter minerals accumulated near the surface. As the Moon cooled, the lighter minerals crystallized according to melting point, then further sank or rose according to density. Radioactive elements were preferentially concentrated near the surface, leading to additional melting or prolonged cooling. All of this resulted in a crustal layer about 50 km thick.

Since the Moon retains no atmosphere, it does not slow incoming objects, and weathering does not occur as on Earth. The original crust which solidified by cooling was later heavily cratered and broken up, but fairly unmodified in composition. The surface is covered by a 2 to 8 meter thick **Regolith** (Lunar soil) made from the original crust plus impacting asteroids. This has been thrown around and mixed during crater formation. The fine regolith is interspersed with larger rocks and boulders (Figure 4.12-3). Because of the broken nature, the regolith is easily collected and moved. Since the layer is global on the Moon, roughly 380 trillion tons of material is available from this source. If this is insufficient, or localized but deeper ores are needed, conventional mining techniques can be used. Blasting is commonly used on Earth for bulk mining, but uses nitrogen-based explosives. That element is in short supply on the Lunar surface. Alternatives include supplying nitrogen compounds or ready explosives from elsewhere, artificial cratering by directing impacts from orbit, small scale fracturing using local rock and accelerators, or **Plasma Torch** cutting using oxygen as the carrier gas, where oxygen is the most abundant element in the Lunar crust. Maximum practical mining depth is limited by overlying rock pressure and crustal temperature gradients, but should be on the order of multiple km. It therefore would be on the order of a thousand times more raw material than the regolith layer.

Industry Survey

We can develop a list of possible Lunar activities by looking at all Earth industry categories, and identifying ones that can potentially operate there in the future. To these we can add any activities that are unique to the Lunar region. Existing industries are classified for statistical purposes by the **North**

Figure 4.12-3 - Large Lunar surface boulder

American Industry Classification System (NAICS). We will use their numbering system and sequence for our survey, and insert unique Lunar items where most appropriate.

11 - Agriculture: Local agriculture will be useful to the extent that people and other living things in the region need food, and it avoids transport from Earth or other space locations. Plants can also recycle waste products from people while producing food and other useful products. Filtered sunlight is widely available in Lunar orbit, but the long Lunar night on the surface, and lack of sunlight underground, may require artificial lighting for higher plants. Microorganisms may be more tolerant of long periods of darkness. Carbon (100-160 ppm), hydrogen (30-60 ppm), nitrogen (60-120 ppm) and water (10-1000 ppm) are rare in Lunar soils, and may need to be imported from other locations (**Fegley and Swindle, 1993**).

21 - Mining: The Moon's surface is covered by an average of 10 million tons/km^2 of already broken up rocks and dust, and 1.36GW/km^2 of daytime solar energy is available to power the mining and processing operations. Mining appears feasible for local use, and to supply locations which take less effort to reach than from the Earth. Delivery of mined products to the Earth itself would be limited. The mineral oxides making up most of the Moon's surface are also widely available on Earth at low cost. There may be some rare materials that will be worth transporting that far, especially if very low cost transport is developed.

22 - Utilities: All activities in the Lunar region will need local power, which at first can be supplied by equipment such as solar panels and batteries brought from Earth. When higher power levels are needed, simple systems like concentrating reflectors and thermal storage can be made locally. As local manufacturing develops, more complex power systems can be built. Our civilization consumes a large and growing amount of energy. 4-15 times as much solar energy is available in the High Orbit region as locations on the Earth's surface, and it is not interrupted by night and weather. If that energy can be delivered at a low enough cost, there would be a large demand for it. The Lunar region is likely too far to efficiently generate and transmit energy to Earth, but can be a source of materials and products to do so in the High Orbit region. The available quantity of solar energy within reasonable transmission range (80,000 km orbit radius) is 27.4 million TW, or over a million times Earth's current total energy use.

23 - Construction: The Earth is large enough to support its human population from a purely physical space standpoint for the foreseeable future. At the density of Manhattan, a population of 10 billion would fit on 0.25% of the world's land area. Even though more space is not necessary, it would be desirable for a number of reasons. The Lunar region can supply some of that physical space. Activities in the Lunar region that involve people and other living things will require habitats to support them, since the natural environment cannot. Large orbital habitats and industrial facilities would be difficult to deliver in a single piece. Either on-site construction is needed, or the orbital equivalent of shipyards and tow ships. Large surface habitats and other facilities would be too difficult to move, and need local construction.

31-33 - Manufacturing: The range of possible manufactured products in the Lunar region is vast. The first to be made are likely to be those that use local resources, displace significant mass otherwise brought from Earth, and are fairly simple. Examples include bulk shielding, thermal storage mass for nighttime power, basic chemicals for propulsion and life support, and mineral products and metals for construction. Once starter sets of production equipment are delivered to the region, they can begin to bootstrap their own expansion, and increase the range of products made locally. Transport energies to and from nearby regions are low, so a wider range of material inputs and product exports for local manufacturing may be feasible.

42 - Wholesale Trade: Individual operations in the region will trade with each other and with other regions, according to the economic principle of comparative advantage. The Moon has large amounts of easily extracted raw materials. So those materials, or products made from them, are likely to flow to Lunar and Earth orbits where they are lacking, and where transport energy is lower than from Earth. Ownership and control of trade may be from Earth rather than local. Communicating with the ground is fairly easy from this region.

44-45 - Retail Trade: We do not expect significant retail trade to develop in the region for a considerable time. Early populations in the Lunar region will be there for scientific and industry purposes. Their employers will likely provide for their basic needs, and even budget for some optional items like entertainment media. As permanent habitation develops and a non-working population accumulates, there will be more time and resources that can be applied to personal choices, and non-essential specialties. Retail trade may then develop.

48-49 - Transportation and Warehousing Transport is required from previously developed regions in order to start any activity in this region. Such transport is accounted for in earlier phases from their starting locations on Earth or in Earth Orbits. Transport within the Lunar region and to farther destinations is accounted for in this phase. At first the transport equipment is all made elsewhere, but over time propellants, structures, and more complex items can be made locally. Transport includes infrastructure such as spaceports and surface vehicle hangars. Warehousing includes all types of storage, which will start with bulk items like unprocessed rock and basic chemicals.

51 - Information Transmitting information through space requires no mass and little energy. So many satellite systems already exist in Earth orbit purely for communications, and nearly all have communications systems to support their primary purpose. Communications networks will be extended to the Lunar region once significant operations begin there. Networking, communications, and information technology for internal use within the Lunar region, and to Earth and other regions, will be extensive from the start. The equipment used is complex and fairly low mass. So it is likely to continue to come from Earth.

52 - Finance and Insurance This will most likely be supplied from Earth, since ownership rights, contracts, and money are all non-material relationships and can be transacted remotely.

53 - Real Estate, Rental, and Leases The **Outer Space Treaty** prohibits claiming celestial bodies but allows peaceful uses in space. International agreements for the Space Station, and orbital slots and frequencies for communications satellites have set precedents for ownership and use of equipment in space. What is not settled is territorial rights to less than entire bodies. For example, if someone builds a mining operation on the Moon, or a colony in Lunar orbit, will they own a territory around it in the sense of land rights on Earth? A reasonable answer is exclusion of others for technical and safety reasons. Thus, someone else cannot build so as to shadow your solar collectors or damage your equipment with rocket exhaust. Until the legal questions are settled, we cannot say what the scope of this industry category will be. Like Finance and Insurance, whatever the scope is, it will likely be handled from Earth for similar reasons.

54 - Professional, Scientific, and Technical: Some research and development, especially scientific, is best carried out locally in the Lunar region. Such activity began in the 1960's, to a large degree for political reasons, and continues to the present. Scientific and increasingly commercial activity is expected to continue in the future in the Lunar region. Most of the people involved in space activity so far have been on Earth, building the equipment and operating it remotely. This is likely to continue for some time, but increasing numbers of people will visit and work in the Lunar region. They will be concentrated in scientific and technical fields at first.

55-56 - Management and Organizational Support: Business management and administration will likely be mostly handled from Earth. Support activities like orbital debris collection and radiation remediation are tasks for Earth orbits.These in turn may be supplied from the Lunar region.

61 - Education: This will likely be done in advance or remotely from Earth at first. Local education will be mostly limited to training for industrial operations. General childhood and higher education will be deferred until permanent habitation develops with populations of young people to teach.

62 - Health and Social Services: Health monitoring and first aid capacity are needed for people in the Lunar region from the start. At first, this will be by training the crews themselves, with remote monitoring and advice. As the number of people grows in the region, more dedicated equipment and health specialists can be supplied. Telepresence, artificial reality, and haptic robots are not currently good enough to do health care remotely, and may not ever be so from Earth, due to speed of light delays. Automated health care carried out locally may be possible. These depend on future R&D, so for now health care concepts would involve existing methods. Nursing and residential care is assumed to be provided on Earth at first, by returning people there. As population grows in the region, local facilities may be established. We assume social assistance will be done remotely or not be needed. Basics like habitation and food need to be provided to everyone in the region for them to live at all. Return to Earth is an option for anyone who needs social support.

71 - Arts, Entertainment, and Recreation: Entertainment will start out as remote delivery or software for people in the region, because their energy and mass are low. Active recreation would start with necessary exercise for health maintenance. Early Lunar landing sites may be preserved as historical locations, and unique sports may develop in the Lunar region, such as rover racing or low-g gymnastics, but these are speculative at present.

72 - Accommodations and Food: Living space, food, and drink must always be provided for people in the region. At first these will be provided by sponsoring organizations and self-operated. As local capacity grows and people establish long-term residences, there will be room for temporary travel accommodations, tourism, and specialty food and drink locations. Space tourism to Low Orbit has already happened, and proposed to the Lunar region, but it is a very limited market due to extreme cost. It does demonstrate there is an interest. The existence of the **Adventure Travel** market makes it likely there will be more tourists once costs become more reasonable.

81 - Other Services: This category covers miscellaneous activities not covered elsewhere. Repair and maintenance is highly desired for the Lunar region from the start. This is because the equipment is either expensive to replace or life-critical. Personal services will start out self-provided until populations are larger. Private and civic organizations are not needed at first, can be extensions of Earth organizations, or self-organized locally.

92 - Public Administration: At first, most of this industry category will be handled from Earth. Fire and public safety will start out self-provided, and develop as specialties as populations grow. Environment quality and monitoring would be designed into Lunar systems as necessities. Publicly funded civil and national security space activities have been the only ones carried out so far in the region, and they are likely to continue to be important. Government budgets for these activities are finite, while business and private ones are not limited in this way, so we expect the latter to dominate eventually. For the sake of efficiency, public and private projects should be integrated so as to support each other

Project Drivers

Which of the above activities make sense for the Lunar region, and when, is a function of a number of factors. They include project motivations, their economics, available technology, and prior and parallel development. These drivers will change over time, affecting which of these activities get started and when.

Motivations

Motivations to develop the region can be personal, organizational, or social. Curiosity drives scientific exploration of the Moon itself, and support of other science in the future. Safety from natural and human-created hazards is another motivation. It has driven orbital weather and defense systems in the past, and may drive asteroid deflection and solar blocking systems in the future. The natural Lunar environment is devoid of life and full of radiation. So moving hazardous activities off the Earth is another safety-driven motivation. The profit motive is ever-present in the business world. The unclaimed material and energy resources in the region are a potential source of profitable activity. [add list of human motivations and those that apply]

Economics

What projects are feasible, and when, is partly driven by available technology and support from previous phases. But it is also driven by the economics of the projects themselves. Economic principles like comparative advantage and returns on investment will still apply in the Lunar region. The first advantage of the Moon is relative closeness to the Earth, and to Earth Orbits which are already in use. Communications and transport times are shorter than for farther regions. Next is the low energy to reach nearby orbits, due to the Moon's smaller mass and position high in the Earth's gravity field. Transport "downhill" (towards the Earth) is eased by slow aerobraking with the Earth's atmosphere. Slow aerobraking uses multiple passes to lower heating and avoid damage. Transport "uphill" (away from the Earth-Moon system) is eased by using the Moon for gravity assists, when both departing and arriving. Lastly, large amounts of raw materials and solar energy are available in the Lunar region to enable projects.

While large amounts of materials are available in the Lunar region, they are not a complete source for all possible needs. The Lunar surface is deficient in volatile compounds, which escaped, and denser minerals and native iron, which sank to the interior. A fully developed economy would need to supplement local materials from elsewhere, driving trade. This includes Near Earth Asteroids, and from lower orbits or Earth. Near Earth Asteroids are a good choice for bulk materials because of the low energy to reach and return from them. Lower orbits and Earth are the likely source for higher value and finished goods, because of the more established industries there. In the long term, other sources may end up being economic. Costs tends to increase with the energy consumed in a task. So trade can make sense when the transport energy from elsewhere is less than the mining and processing energy from scarce local sources. Trade, in turn, drives a need for low cost transport.

Earth already maintains a complete civilization. So physical products from the Lunar region are not likely to compete with Earthly ones on cost. Exceptions include scientific and collectible samples, where their value specifically comes from their Lunar origin. Information is low mass and fairly easy to deliver, so may have significant economic value when coming from the Moon. Energy is a very large market on Earth, and solar energy is much more available in Earth orbit. If it can be delivered economically, that would be a spur to space industry in general, and for exports from the Lunar region to support it.

Technology

As of 2016, launch costs from Earth to the Lunar region were about $55 million/ton. This is somewhat higher than the price of gold/ton, which severely limits what kinds of projects are feasible. The raw wholesale energy to reach the Moon is only $1000/ton, which shows how much room there is for improvement. A number of new transport systems are in development. The SLS may eventually reach a cost of $1 billion/flight for 55 tons to the region using all chemical propulsion, or $18 million/ton. The SLS and an efficient electric tug from low orbit could deliver about 105 tons for $1.2 billion, or $11.5 million/ton. The Falcon Heavy can lift 64 tons to Earth orbit, and with an electric tug could deliver 50 tons to the Lunar region for about $200 million, or $4 million/ton. These are improvements, but still leave costs very high. For comparison, the mid-2017 price for silver was $0.5 million/ton. Any transportation measured in multiples of a precious metal price is still an impediment to significant development.

We therefore have a strong incentive for two kinds of technical improvements: systems with even lower transport costs, and use of local resources to avoid transportation. Both require extensive R&D to get the new technologies ready before attempting large-scale lunar development. Both improvements are useful for other locations. So we expect their development to be started in earlier program phases, and upgraded as needed for Lunar development.

Prior and Parallel Development

The technology improvements noted above will not all happen at once, so different lunar activities and projects will become feasible at different times. This will be in parallel with other program phases on Earth and in space, and other development in the Lunar region either before or in parallel. To the extent that transport costs to all parts of space are still high, it favors high value/low mass activities, and those where economic returns do not matter, such as publicly funded research and exploration. Low and high Earth orbits will likely continue to be more developed than the Moon, and therefore drive the improvements in transport and resource use. High thrust systems may be adapted from other regions for initial Lunar surface access. Improvements like Skyhooks, which are initially developed for Earth orbits, can later be adapted for Lunar surface access. Asteroid processing may start at L2, which is outside the Lunar region, but additional locations near or on the Moon may be added later. These locations can use the same or upgraded equipment.

Development Projects

We can now start to combine the above information into a general approach for Lunar development, and identify specific projects. We can provide initial concepts for these projects, which will gives a sense of their scale and main features. But this is merely a starting point, and does not exclude alternate ideas. It also barely begins optimization and integration of the projects to each other and other projects elsewhere.

General Approach

The Moon's surface is the size of two continents on Earth, and the orbital region has a cross section of 3.85 billion km^2, or 7.54 times the Earth's total surface. This region is far too large to develop all at once, or by a single project or organization. Our general approach is then to identify a number of smaller tasks and projects. They can be individually carried out by different organizations and put into a logical sequence. They would interact with each other when they exist at the same time, with other program phases, and with the rest of civilization.

The tasks can be grouped by start times into preparation, orbital, and surface development. There are already a number of current and near-term projects aimed at the early tasks. We note these next, and their existence should be integrated into plans for later projects. We don't have an integrated sequence yet for specific longer-term projects. So they are listed by primary function (production, habitation, transport, or services) and location (orbital or lunar surface) instead. Within each category we list them them in approximate start time order. We expect earlier projects to overlap later ones in the same category, rather than be completely replaced.

Preparation - Planning and designing future Lunar projects requires understanding the relevant features of the region in general, and of specific operating locations. Preparation for Lunar development therefore involves tasks like exploration, surveys, prospecting, and site investigation. Scientific exploration of the Moon began as soon as telescopes were available to observe it in detail. It accelerated once rocketry enabled placing instruments in orbit and on the surface, and bringing back samples for analysis. The orbital region is now well enough understood to begin projects there. For the surface, surveys and prospecting of the Moon as a whole are ongoing, but detailed work for future surface activities is incomplete. We assume devices like surface rovers will be sent before substantial development of a particular location. They would do detailed surveys, prospecting, and subsurface investigation.

Orbital Development - Lunar orbit is easier to reach than the surface, especially at first. So we assume orbital development begins earlier than the surface. Orbital locations generally don't need physical preparation, as they are devoid of natural materials. Use of a given orbit can thus begin with delivery of equipment. Smaller satellites, such as for communications relay, can be delivered as a complete unit from Earth. Larger installations like an inhabited station can use sections produced on Earth, and assembled in Earth orbit or in their final location. In either case, an electric tug can deliver it slowly but efficiently. People would first travel by faster chemical rockets to avoid radiation exposure. The same electric tugs can fetch asteroid rock from nearby orbits. It can first be used as-is for bulk radiation shielding. Processing equipment brought from Earth can then start making simpler products, like propellants, oxygen, and water, out of the asteroid materials.

We don't yet know the best path for further development. A lot more R&D is needed on this subject, but for example we can follow history and use iron and steam to bootstrap a growing economy. Metallic asteroids contain mostly iron, alloyed with varying amounts of nickel and cobalt. Chondrite asteroids have up to 20% water and carbon. The carbon can be added to the iron to make steel, and the water heated with focused sunlight for steam power. A starter set of production tools would convert the metal stock to parts for power units and more production machines, bootstrapping their own expansion. They can also make structural parts, pressure vessels, and other items to widen the range of production processes. This may be a simpler way to bootstrap than producing solar panels at first. Later mining and transport from the Lunar surface would increase the quantity and diversity of source materials. Rare materials and hard-to-make items will continue to come from Earth, but a decreasing percentage as production builds up. New people, remote control of Lunar systems, and support services are still be provided from Earth.

Surface Development - Lunar surface work starts later than in orbit, but continues in parallel afterwards. Sites on the surface are typically not ideal in their natural state. So construction sitework can begin after prospecting and site investigation. This would use some combination of robotic, automated, and remote controlled machines delivered from Earth. Propellants produced in orbit make their delivery easier. Some kinds of projects don't need large numbers of people. They can continue with remote operation, and perhaps short visits by crew for maintenance. Short visits and smaller long-term populations can feasibly be supported by deliveries from elsewhere. Delivering everything for larger populations and industrial operations is inefficient. Local mining, processing, and fabrication would be built up over time to support them.

Like for orbit, the best path for growth on the surface needs extensive R&D. It could begin with delivery of ready production equipment, then use starter sets to bootstrap further expansion. Depending on size and cost, local production of simple products may be started as early as the remote operation stage. These can be stockpiled for later development. Even after people are supported part or full time on the surface, you can continue to use machines to augment their work, using self- or remote-control. Depending on the level of automation, there may be high numbers of such machines per person. This approach will leverage the limited early human population.

Current and Near-Term Projects:

Transport from Earth - Direct transport from Earth is being worked on by NASA through their **Space Launch System (SLS)** and **Orion Spacecraft** projects. The SLS provides the capacity to deliver large payloads to the Lunar region (and other destinations), and Orion for carrying people. **Other Chemical Rockets**, which are in varying stages of operation and development, can also deliver payloads to the region. Only the larger ones can deliver significant payloads unaided. The **Scaled Composites Stratolaunch** carrier aircraft is part of a hybrid turbofan/chemical rocket transport system and can potentially support Lunar missions. The combined cycle **SABRE** airbreathing/rocket engine for the **Reaction Engines Skylon** spaceplane is in early stages of development. Skylon is intended to carry 15 tons to low orbit, enough support Lunar missions.

Transport from Low Orbit - Chemical rocket performance is limited by the available energy of the propellant. It is used for the Earth to Low Orbit transport segment because it provides high enough thrust to overcome the Earth's gravity and prevent the trajectory from intersecting the surface again before reaching orbit. Once Low Orbit projects are built in Phase 4A, slower but much higher performance alternatives can be used, and transport infrastructure built up from Low Orbit to the Lunar region. Even with chemical propulsion, system cost can be lowered compared to the current practice of single-use transport. This is possible if a cheaper propellant source is available or the hardware can be used multiple times. Multi-use orbital transport vehicles are generally referred to as **Space Tugs** by analogy to tugboats, which push ships and barges around.

Missions to the Lunar Region - There have been 64 successful public **Missions to the Moon** or involving the Lunar region since 1959. Notable among them were the Apollo missions which carried people and brought back 382 kg of samples. About 20 more public and private probes are under development, intended to launch between 2017 and the 2020s. Two short-term crewed missions are also in development. A dozen more probes and crewed missions are currently proposed but not yet in development. The NASA **Deep Space Gateway** is a crew-tended station planned for the 2020's in a halo orbit near the Moon. The European and Chinese Space Agencies have proposed Lunar surface bases by the 2030s. Both may involve international and private participation. We expect public projects in the region to continue to evolve, but be limited by available funding.

Long-Term Projects:

As noted in the general approach above, a lot of research and development is needed before we can make definite plans for long term Lunar development. What we can do now is identify potential projects and what work is needed to prepare for them. They can be put together into a plan, but this should be considered preliminary and very likely to be revised as time goes by. Projects are grouped by function and location for convenience, then in approximate time order.

Lunar Orbit Production

Lunar orbits require less energy to reach than the surface, and are easy to reach from the High Earth Orbit region because the Lunar region is embedded in it. So we expect orbital production to precede surface production. Early production can be an outgrowth of public projects such as the Deep Space Gateway. These projects will begin with assembly of pre-made components brought from Earth. Local production can then be added incrementally. Production can start with pre-made tools and equipment, then bootstrap its own expansion by using the seed factory approach. A starter set of equipment is used to make some parts for its own expansion, with the remainder imported. Over time, more can be made locally, to the limit of what makes economic sense. Production outputs can be used locally in orbit, delivered to the surface, exported to existing markets in the High and Low Orbit regions, and exported to farther destinations. We do not expect much export to Earth from the region because civilization is well developed there and costs are likely lower. Major production steps include:

Materials and Energy Sources - There are no significant materials sources in Lunar orbit, so they must be imported. Despite greater physical distance, Near Earth Asteroids are the easiest source of materials at first. This is because electric propulsion can be used, which is very efficient, and the Moon can be used for gravity assists, reducing propulsive delta-V. Neither is available for early transport to and from the surface. Asteroids also have more varied compositions than the Lunar surface, allowing a wider range of products. When transport infrastructure such as catapults and skyhooks become available (see Lunar Transport below), the lower cost, higher volume, and shorter delivery time from the Lunar surface becomes an advantage, shifting the balance to more local materials supply. Some materials are rare in both asteroid and Lunar ores, and would still need to be supplied from Earth.

Orbital locations have solar energy available 50-100% of the time, mostly as a function of altitude, while surface locations have it 50% or less, decreasing with latitude. Since the embodied energy to make products is typically larger than the kinetic energy to move materials from the Lunar surface to orbit, production can generally be faster in orbit. Most of the markets for orbital production will be local or other orbital regions, and require less energy for delivery than from the surface. Orbital production should continue to be favored for these reasons.

Materials Processing - Some products, like bulk radiation shielding, don't require further processing, just delivery. But most require conversion of raw materials to finished stock by mechanical, thermal, chemical, electrical, or other methods. Processing can start with the easiest and simplest methods. An example is water extraction from Chondrite type asteroid rock, which only requires heating a container using concentrated sunlight, and a shadowed condenser to collect the evolved vapor. Metallic asteroids are dominated by native iron-nickel-cobalt alloy, which makes up more than 95% of their mass. However individual samples vary in composition, and may have undesirable trace elements, lack other desirable alloying elements, or contain rocky inclusions. The Stony-iron group have higher percentages of silicates and less native metal. Processing is therefore generally needed to produce uniform alloys in desired sizes and shapes. The stony fraction of stony-irons, and the pure **Stony or S-Type** asteroids are typically silicate minerals which contain various metallic elements. These require extensive processing to extract desired products.

Orbital Fabrication - There are many known fabrication processes to convert stock materials into finished parts. These include all the known methods used on Earth. To these we can add some unique methods which take advantage of zero gravity, vacuum, and full-spectrum solar energy. Which ones are suitable for Lunar orbit and in what sequence to develop them will requires a lot of R&D. At a minimum, processes known to work on Earth can be used unchanged in space by providing a normal atmosphere and artificial gravity. However, these may be quite massive and inefficient in orbit.

Orbital Assembly and Construction - This can begin with prefabricated elements delivered from Earth and assembled in orbit, such as space stations and larger and more powerful satellites. As local production develops, the orbital products are less constrained by launch mass and cost from Earth, so they can use simpler and heavier designs. The Lunar orbit region is a useful location to combine raw materials supply, energy supply, and parts delivery from Earth or lower orbits. It may therefore develop as a major assembly and construction site. Large projects like space colonies may be too massive to move once constructed. In that case, smaller elements may be produced in Lunar orbit, then delivered to their final location for construction and installation.

Production systems on the surface can make some items without using imported materials. The relatively limited variation in Lunar ores means a wider range of products will need significant outside supplies. Landing on the Moon requires significant propellant at first, so imported supplies have a higher cost. So early production will favor those items that can be mostly or entirely made locally. Once efficient two-way transport is available (see Transport below), large amounts of raw materials and unfinished goods can be exported for orbital production. More finished goods and materials can also be imported for surface production. This allows a wider range of items to be made on the surface, and results in a trading network that benefits both locations. The preferred locations for surface production depends on availability of energy and material sources, local conditions, transport capacity, and the intended destinations and uses for the products. For example, a scientific outpost may not want industrial operations nearby if they cause unwanted disturbances. Possible products include:

Minimally Processed Regolith - Early Lunar mining and construction would be in support of near-term public scientific and exploration projects such as a surface station. This would include clearing and leveling building sites and access paths; then excavation, placement, and covering of station elements with unprocessed local regolith for radiation, thermal, and debris protection. Debris protection is both from natural meteoroids, and materials thrown by lander exhaust. Covering may be by simple loose piling of material over a support structure, and not require processing beyond sorting for rock size. Paving and blocks may be produced by melting the soil with concentrated sunlight.

Water Extraction - The Moon is devoid of known liquid water, but **Lunar Water** is known to exist in two main forms. The first is chemically bound as hydrates and hydroxide minerals. High temperatures are required to extract the water from this source. The concentrations are ~10-1000 parts per million. The second is water ice trapped in permanently shadowed craters near the poles, where temperatures are low enough to preserve it. Asteroids can contain up to 20% water, and Hydrogen can be combined with abundant mineral oxygen to produce 9 times its mass in water. So import is an alternative to the low concentrations and limited locations of local sources. In the long term, water is abundant beyond the mid-Asteroid Belt **Frost Line**, where temperatures have stayed low enough to retain solid ice. If low cost transport is available, they would be a preferred source over the limited Lunar sources. Of course, Earth has a great deal of water, but transporting it requires a lot of energy.

Native Iron Production - An early production path may be to mine and process native Lunar iron, which is present at about 0.5% of Lunar soil. If this process works, it may be feasible with fairly simple and low mass equipment, giving an early return on hardware delivered to the Moon. The soil is sieved for small grains, and iron-bearing particles are selected by magnetic separation. The remainder are used to create sand molds. A robot smooths a patch of ground, spreads the non-iron grains over them, then compacts and makes depressions for molds. The iron-bearing particles are placed in the mold cavities, and a large concentrating dish focuses sunlight on the cavities, scanning them sequentially if needed. The iron particles are often attached to other mineral grains, but being denser it will sink to the bottom of the mold, and glassy slag will rise to the top. The result would be basic shapes like plates and bars, with slag stuck to it on both sides from melting.

Depending on the melting point of iron vs other minerals, it may be necessary to select for refractory ones for the molds. One way to do this is to heat samples until the less refractory components sublimate. Another is to prospect for suitable refractory source rocks. A third approach is to bring a highly refractory crucible in which the grains are melted, and pouring off the slag and iron separately. In that case the mold is hot for a short period and will not melt much. R&D is needed to determine which, if any of these approaches will work. The cast metal stock will need to be sand-blasted to clean off adhered slag and mold grains. Fortunately there is no shortage of mineral grains on the Lunar Surface to do that with. You now have an inventory of iron stock to use for all types of construction and parts making.

Chemical Processing and Metallurgy - More complex processing may require more equipment, but yield a variety of stock materials from a given amount of Lunar ore. There is generally about 43% oxygen in any given soil, so any of several reduction processes can be used to extract it. Oxygen is useful as a propellant, and for life support. The soil averages about 21% silicon, which can be used for solar panels and in steel for transformers and motors. Combined with oxygen it makes quartz, and with other elements makes common glass for mirrors and windows. The soil averages about 24% iron, aluminum, magnesium, and titanium, which are useful structural metals. Lunar basalt, which makes up most of the maria regions, can make high strength basalt fiber. So with full processing, nearly all of the soil can be turned into useful products. However, the soil is low in certain elements like carbon, hydrogen, and nitrogen. The first is used in making steel, and all three are needed for anything organic. These may be imported rather than extracted from low-grade ores. Early markets will likely be to previously developed regions, and more solar energy is available for production in higher Lunar orbits. So we expect early production will mostly be for export to Lunar orbit and beyond, with production for local use gradually increasing.

Helium-3 Mining - Mining of **Helium-3** from the Moon has often been suggested for terrestrial power, because it has low radiation by-products in fusion reactions. There are several problems with this idea. The first is that Deuterium/Helium-3 fusion is about ten times harder than Deuterium/Tritium. We don't yet know how to sustain D-T fusion, much less do it economically, and we are even farther from doing the harder D-He^3 fusion. So we simply don't need it right now. The second is He^3 is implanted on the Lunar surface from the Solar wind. However this source is diffuse, and this isotope is extremely volatile. The net abundance is therefore only 3-15 parts per billion on most of the Moon. We would therefore have to process a billion tons of Lunar soil to get 3-15 tons of product.

The abundance is 1000-6000 times higher in outer gas giants because their atmospheres contain 15% (Uranus) and 19% (Neptune) Helium, and therefore proportionally more of the He^3 isotope. The higher ore concentration requires proportionally less tonnage of mining and processing to extract a ton of product. Although the outer planets are much farther than the Moon, if we need He^3, then D-T fusion would already be solved, and it can power ships to reach to reach those planets. Since those atmospheres are mostly hydrogen, the Deuterium isotope is widely available. Tritium can be made from He^3 by neutron bombardment. The bulk hydrogen and helium from the atmospheres can be used as propellants. Thus mining the outer planets can be self-sustaining. The third problem is relative energy content. Pure He^3 can supply 200 TJ/kg, but at Lunar concentrations the mined soil can only supply up to 3 MJ/kg in reaction energy. This is only about a tenth the energy in fossil fuels and a fifth that in wood. So it would hardly be worth mining on Earth, much less the Moon.

Silicon and aluminum make up 28% of typical Lunar soil. Assume only 10% of the soil mass is extracted to these elements and made into solar cells and structure for power satellites. The satellites will produce about 100 W/kg, so the net power per mined kg of soil is 10 W/kg. A nominal life for such cells is 15 years in orbit, during which they would produce 4.75 GJ/kg of energy. This energy can be beamed to Earth around the clock. So if you wish to mine the Moon for terrestrial energy supply, you can generate 1500 times as much per kg via solar power than He^3 fusion, even at low elemental extraction efficiency. Space solar cells degrade due to radiation damage. However the damaged cells represent a high quality ore for reprocessing. It should take less energy to do this than the original extraction of silicon from rock, and the cells already produce much more energy than used to produce them. So in principle we can extend the life of the satellites indefinitely. In that case their energy production is only limited by the life of the Sun.

As noted above under Available Resources, some regions of the Lunar surface contain up to 10 parts per million Uranium and Thorium. These elements were concentrated in the crust by differential processes as the

Moon cooled from a molten state. The mined ore has an energy content of about 800 MJ/kg, or hundreds of times the He-3 energy content. So even for nuclear power on the Moon or other space locations, He^3 is not the most efficient source. Solar is even higher energy content, but there are situations where nuclear power is useful. These include spanning the long Lunar night, in shadowed craters, or for portable power, and for destinations far from the Sun where solar power is weak. The conclusion is mining He^3 from the Moon does not make sense for any purpose we can see.

Surface Fabrication and Assembly - Like for other program phases, we use a bootstrapping process to build up our production capacity. We begin by importing ready-to-use tools and equipment, such as a mobile solar furnace for paving and block-making. We then add a starter set of machines (seed factory) capable of fabricating parts for additional production machines, as well as other end products. Machine tools such as lathes and milling machines are designed to make metal parts from metal stock. Modern ones are computer controlled, and are themselves mostly made out of metal. So an initial set of machines, when supplied with local stock materials, can make many of the parts for more machines. Given plans for a wide variety of machines, the initial set can increase in both size in complexity. For example a forging press or rolling mill are useful for making some types of metal parts, but those are likely too heavy to bring to the Moon to start with. But if you have a good supply of local materials, you can build them later when needed.

The starter set cannot make all types of needed parts, so the remainder continue to be imported. As the collection of machines grow, imported parts and materials decrease as a percentage, though they may increase in absolute amount. Parts are then assembled using some combination of robotics, remote-control, and direct human labor. Early assembly may be as simple as stacking blocks for a radiation shelter, but more complex products will need dedicated assembly space. Robots and factory buildings are themselves products which can be fabricated and assembled, so the entire production system can grow itself over time.

Electric Power - Energy is needed to operate all kinds of production on the Lunar surface. This includes electrical power and process heat. Early electric power systems can be brought as complete systems from orbit. This may be photovoltaic solar panels or nuclear. Mass is not an issue for a stationary facility, so solar-thermal using local materials is an expansion option. Complex equipment such as a steam turbine-generator set can be delivered. Reflectors to provide concentrated sunlight can be made locally. These can be sheet metal with an evaporated aluminum coating to make them more reflective, or in a simpler version pre-made reflective sheets are delivered and only the support structure is made locally. The percentage made locally can increase over time as the range of production processes grows. Water and rock from local sources supply the working materials. A container is filled with Lunar rock and a little gas for heat transfer. This is surrounded by sized regolith particles for vacuum insulation. During the day the container is heated with concentrated sunlight, and the heat transfer gas directly used to boil water and run the generator. At night, the stored heat in the rocks continues to heat the gas. Large amounts of rock are available on the Surface, so we can provide two weeks worth of storage, and a crater can supply a convenient place to put the container and insulation.

Process Heating - Many industrial processes need heating in some form. This can also be produced with concentrated sunlight the same way as for electric power, but substituting a suitable heater or furnace at the focus rather than a thermal storage device. Different processes need different amounts of heat, and the temperature for a given process can vary over time. This can be provided with an array of steerable mirrors directed to chosen targets as needed, and a blocking shutter for fine control. The furnaces and heaters would likely want to be stationary for a number of reasons. Since the Sun moves in the sky, the mirrors have to be steerable anyway, so directing them to different targets adds little complexity to the design.

Lunar Orbit Habitation

[TBD]

Lunar Surface Habitation

Habitation in general is the facilities that allow people to occupy a given place. On Earth this includes homes, and the parts of commercial and industrial facilities designed for people's needs. For example, rest rooms and break areas in a factory are designed for people, while the production machines are designed for whatever products the factory makes. Since the Lunar surface is hostile to human life, habitats there must protect from the natural environment and provide all of people's basic needs. This includes air, water, food, temperature control, sleep, sanitation, and others.

Near Term Habitats - Habitats for near-term science and exploration missions to the Lunar surface will likely be based on pressurized modules, like the current International Space Station uses. They would be pre-fabricated and delivered from Earth, then assembled on-site. This will require equipment to prepare the site and move the pieces from where they land to their final location. These may be either rigid-shell or inflatable design. Modules and other equipment may need protection from daily temperature variations, radiation, meteorites, mechanical and electrostatic dust transport, and human-caused hazards such as lander rocket exhaust. A general approach is to use arched structures similar to ones used on

Figure 4.12-4 - Corrugated steel building on Earth.

Earth (Figure 4.12-4). These are covered by Lunar soil and rocks, either in natural form or formed into blocks. The structure allows access to the outside of modules for maintenance and additional storage and work space for movable items. Since low gravity is known to be harmful over extended periods, crew time would be limited to a year or less. For small numbers of people it is feasible to supply them from orbit.

Long Term Habitats - Habitats involving longer stays or permanent residence, and larger numbers of people, will need different designs. People can withstand cramped quarters for short times, but for psychological and personal reasons they will want more space for long-term stays. To maintain health they may need centrifuges to create artificial gravity. An example design would be a large habitat dome for spaciousness, and a centrifuge built around the rim for living or exercise needs. Residents would spend enough time in the centrifuge to maintain health, but could work and enjoy the low gravity the rest of the time. We have essentially no data on how much gravity is enough between zero and 1.0. We know the body deteriorates over time in zero gravity. So as a worst case, people would need to spend most of their time in a one gee centrifuge of some type, but this subject needs more research.

Large habitats can be assembled from pre-made sections delivered from elsewhere, or use local production of structural elements. Until sufficient capacity for people is in place on the surface, it makes sense to do much of the work via remote-controlled robots. Stations in Lunar orbit or at one of the nearby Lagrange points would have short communications time. This may allow better real-time control than from Earth. Since the natural environment is hostile to people, safety is a critical design factor. This would include multiple layers and compartments to contain atmosphere, and emergency breathing apparatus in multiple locations. It would also include failsafe or multiple life support systems. Failsafe design would still support residents in case power or equipment fails, though at reduced capacity or duration. Growing plants has been demonstrated in zero gravity, so it may be possible to grow food in natural Lunar gravity.

Research is needed to make sure it is feasible and does not create problems in the food products. We also need to determine if Lunar soil can be used or is even desirable, lighting for the plants with an abnormal day length, and how to supply organics and other plant nutrients.

Lunar Orbit Transport

Lunar orbit transport includes projects and systems based in this orbital region. Ones which are based in previous regions, such as Earth orbits, are assigned to their respective program phases.

Reusable Landers - Early landings on the Moon do not have the support of much infrastructure, so they require high thrust chemical rockets to navigate the Lunar gravity field. For transport down to the surface, Carbonaceous type asteroids contain up to 20% carbon compounds and water. This can be reformed chemically to hydrocarbons and oxygen, which is a common rocket fuel combination. They would be produced at a high orbit location where there is full-time sunlight for power. An electric tug delivers them to low Lunar orbit for efficiency. A propellant depot then fills a reusable lander which delivers people and early cargo to the surface. Oxygen is the most common element on the Moon, and a number of ways of extracting it from oxide minerals have been studied. An oxygen plant to refuel for the return trip can reduce the round-trip mass ratio from 2.9 (LOX/CH_4) to 1.96. This increases payload per trip from ~25% to ~40%.

Orbital Tugs - These are needed for efficient transport within Lunar orbits, and to and from these orbits to other regions. Transport to the High and Low orbit regions can be relatively low energy. In some cases the Moon can be used for gravity assists to lower orbits, and aerobraking used to lower apogee. Gravity assists can also be used to escape from the Earth-Moon system, but upwards transport generally requires more propulsion. Gravity and drag do not help in changing orbits around the Moon, but total velocity requirements are fairly low. We expect most of the propulsion will be electric. Orbital tugs will likely be developed first for Earth orbits, and just copied for Lunar uses.

Lunar Skyhooks - Lunar basalt and carbon from asteroids can be turned into high strength fibers. These can be used for a large rotating structure known as a "Skyhook", where the rotation at the tips cancels part or all of the orbit velocity. For an orbit velocity of 1564 m/s, which corresponds to an altitude of 262 km, a radius of 250 km produces 1.00 gees at the tip while canceling the orbit velocity. The low point of rotation is then 12 km above the surface, which misses any mountains and allows for the irregular gravity field. Vehicles can then arrive and be dropped off at zero horizontal velocity and 12 km +/- the elevation of the landing point above the surface. This will require minimal propulsion to land and take off. 1.00 gees at the tip allows a crew to live comfortably in orbit, and make trips to the Lunar surface with very low fuel use. A "frozen" orbit at 86 degrees inclination minimizes orbit perturbations and allows access to nearly all of the Lunar surface. However it limits the times when a particular location can be accessed. An equatorial orbit would have more frequent access but more limited surface coverage. Unlike Earth, the Moon rotates so slowly that the benefit of its own rotation in reaching orbit is negligible.

An orbital Skyhook only makes economic sense if the traffic to the Lunar surface is large enough to justify it. This will not be true at first, and bulk material from the surface can aid in its construction, so it would not be built right away. There is a one-time cost in mass to build a Skyhook, but afterwards it saves most of the required lander propellant. A Skyhook can be built incrementally, and provide partial savings before completion, which affects the economics. At the other end of the rotation, the tip is moving faster than Lunar escape velocity, and can therefore send vehicles to a large range of orbits by choosing the radius and time of release. Catching and releasing vehicles affects the orbit of the structure, but if traffic is balanced in direction it is a temporary change. Ballast mass can be provided at the center to decrease the orbit changes, and zero-gravity activities may accumulate there in any case. If traffic is more in one direction than the other, the difference can be made up by electric propulsion. Since low gravity is known to be harmful, an orbital centrifuge can allow crew to mostly live in normal gravity, and operate equipment by remote control with short short communication times. At the same time, the centrifuge can provide easy access to the surface when needed. It is not purely a transport system, but more like a spaceport in the sense of an airport on Earth is a hub for surrounding activity.

Continued on page 2 →

Section 4.10: Phase 5B - Mars Development

Mars has been a leading location in popular culture for future colonization. This is because the natural environment is the most Earth-like by a number of measures. Numerous **fictional works** have described the possibilities, with varying amounts of realism. The **Library of Congress** lists 450 works primarily about Mars as an astronomical object, 73 as an objective of space flight, and numerous other non-fiction works reference it, such as those on planetary science. As of 2017, 25 successful **Missions** have flown past or arrived at Mars, along with a number of failed attempts. Eight of these are operational, and seven more are in various stages of development. Human missions are a goal for NASA, and colonization is the stated goal of the private company SpaceX.

However, most of this attention and work has focused on Mars alone, to the exclusion of the resources and development of the rest of the Solar System. The program we describe in this book is more inclusive. It treats Mars as one place among many to develop. This is partly because other places have different material and energy resources, and their local environments are more suited to some purposes than others. It is also because people are different, and don't all want to live or work in the same kinds of places.

As in previous phases, we would bring starter sets of equipment to the Mars region. They use local resources to bootstrap their own expansion and begin development. This adds Mars to the growing network of developed regions. By the time Mars development is started, this approach would have been used seven times on Earth and in previous space locations, and should be well understood. Leveraging equipment and resources from earlier phases, plus from Mars itself, should enable large scale development of Mars at low cost. This is in comparison to "Flags and Footprints" projects like Apollo. They leave no lasting infrastructure or economic activity, and so each mission requires everything come from Earth at the same cost. The return to cost ratio for our approach should be much higher, and may even be economically self-supporting.

We begin concept exploration for this phase by describing the characteristics of the region and a survey to identify possible future activities. We then look at motivations, economics, and technology, and combine all the information to identify a development approach and set of projects by function, then link them to the other program phases. Projects for which we have additional details and calculations are then more fully described at the end of the section. As an output from this concept exploration, we identify what R&D will be needed for this phase, and feed it back to the preceding Phase 0N - R&D for Mars Development.

Mars Region Features

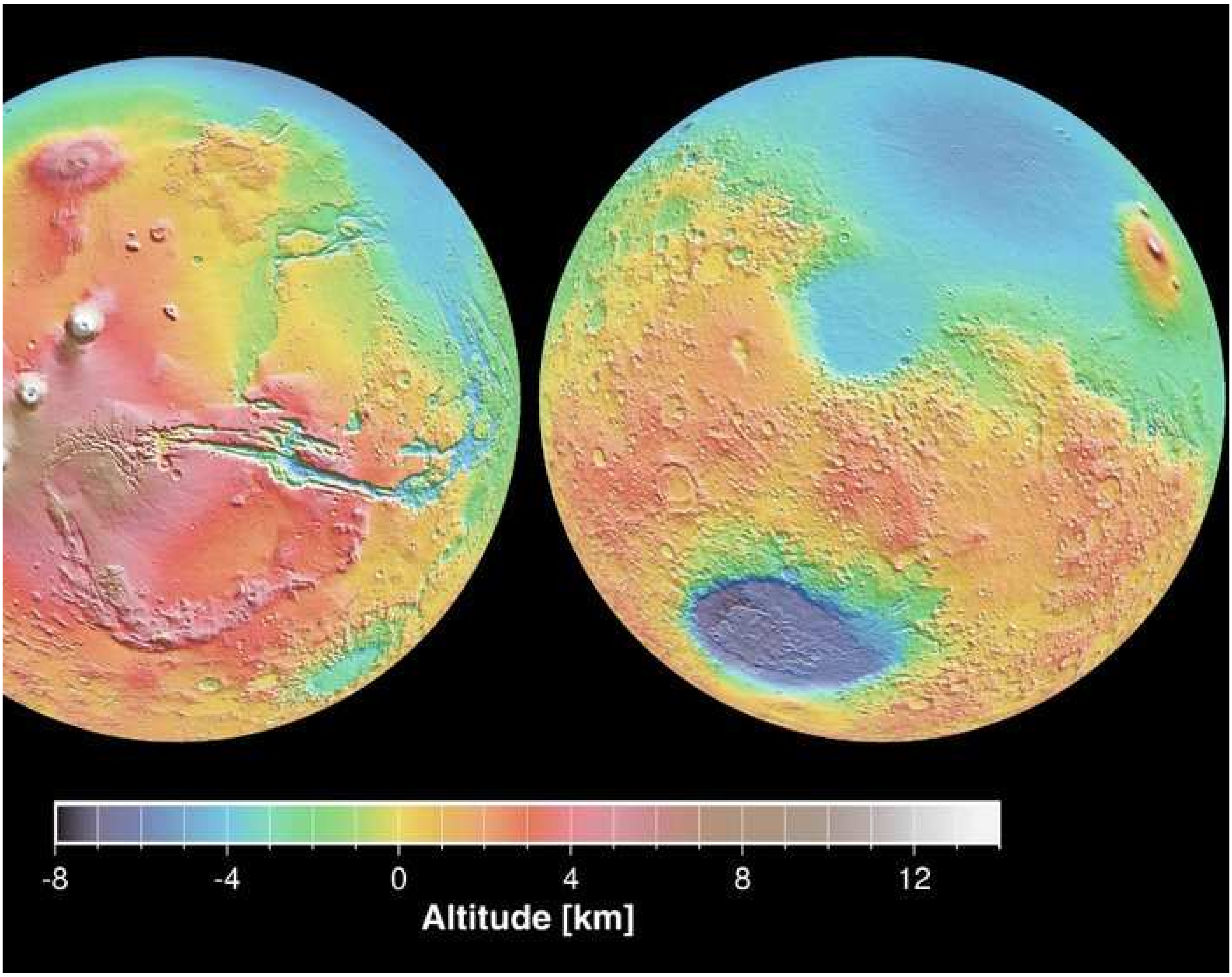
The Mars region is embedded in the Inner Interplanetary region, which extends from near the Sun to a distance of 1.8 AU, and encompasses the regions around the four inner planets. This region includes the planet, two small moons, and orbits within 340,000 km (100 radii) of Mars' center. The Martian moons **Phobos** and **Deimos** have orbit radii of 9,376 and 23,460 km respectively, placing them well within the region. Their orbits are nearly circular and equatorial to Mars.

The Sun is 3,098,000 times Mars' mass, and at the edge of the region averages 670 times the distance. Therefore the Sun's gravity is 6.9 times stronger than the planet's at this distance. However, orbital stability depends on the cube root of the mass ratio, because it is the difference in solar attraction on the planet vs. on points along orbits which disturbs them. So orbits within 1,084,000 km of Mars are at least theoretically stable. We set the region boundary somewhat arbitrarily at about one-third of this, where objects are within 1% of escape energy.

The horizontal surface area of Mars is 144.8 million km^2, which by coincidence is 97% of Earth's land area, or nearly equal. Elevations relative to a reference height range from -8.2 km in the Hellas Basin to +21.2 km on Olympus Mons (Figure 4.14-1). See **USGS Topographic Map of Mars, 2013** for a more detailed version. The Martian day length is 24h 40m, slightly longer than Earth, and the Martian year is 1.88 Earth years. Both moon's rotations are synchronous with their orbits, keeping the same side facing the planet. Their day lengths are equal to their orbit periods of 7.65 and 30.31 hours. The planet is tilted 25.2 degrees to the orbit plane, resulting in seasonal changes similar to Earth's from a geographic standpoint. They are larger in amplitude due to an orbit which varies from 1.381 to 1.666 AU from the Sun.

Environment Parameters

Temperature - The **Climate of Mars** has important similarities to Earth's, including polar ice caps, seasons, and weather patterns. Due to greater average distance from the Sun, surface temperatures vary from 120 to 293K (-153 to 20C), between the poles and noontime at the Equator. Average temperatures change by about 1-2 K/km of elevation. Typical day-night variation is 70 K because the thin atmosphere does not have much thermal mass. This variation is reduced to about 10 K during dust storms. The surface is mostly covered in sand and dust, which moderates variations



opographic map of Mars.

towards the mean annual temperature as depths reach 1 meter. Subsurface temperatures are poorly known, due to a lack of direct measurements. Models based on indirect data indicate they may increase 6-10 K/km of depth, with the lower values in ice-saturated ground.

Temperatures in orbit and on the moons will depend mainly on percentage exposure to direct sunlight. This is lower for low orbits, which spend more time in the planet's shadow. The moons keep one face towards Mars and their shapes are irregular. So available sunlight, and thus temperature, will vary by particular surface site. Mars serves as a secondary infrared source which fills almost 50% of the view in low orbits, and only 0.00275% at the edge of the region. Black surfaces in full sunlight will have equilibrium temperatures of 302-332 K (29-59 C) depending on where Mars is in its orbit. Actual hardware temperatures in orbit will depend on their solar exposure times and angles, albedos, emissivities, and thermal properties.

Atmosphere and Water - Mars has an **Atmosphere** which is 96% CO_2, a bit under 2% each Argon and Nitrogen, and an assortment of trace gases. It extends about 200 km vertically to the point the exosphere merges with space vacuum. Surface pressure varies from 30 Pascals at the top of Olympus Mons to 1155 Pascals in the Hellas Planitia basin. For comparison, the highest value is 1.14% of sea-level pressure on Earth. Pressure decreases exponentially with elevation, with a scale height of about 11 km per factor of ***e*** change vs. a 636 Pascal reference value at zero elevation. The pressure varies by 30% annually with distance from the Sun, as some of the CO_2 freezes and evaporates at the poles. The moons are too small to retain any atmosphere.

The planet has remnant magnetic fields in some areas, which are about 100 times lower strength than Earth's, but no core-driven global field. Therefore it does not have a strong magnetosphere. The interplanetary solar wind dominates most of the region's space environment. The planet creates a **bow shock** on the side facing the wind, a magnetic pile-up of ions, and a rarefied downstream wake and tail, through which some ionosphere material leaks.

Mars has significant amounts of water in the soil as hydrated minerals and permafrost, and frozen in thick dusty ice caps. The ice caps contain about 1.75 million km^3 of ice, and total **Water on Mars** may be 5 million km^3. Water content in the top meter of ground is generally high above 65 degrees latitude, but is still a few percent even at lower latitudes. Water content at depth on the planet and for the moons is poorly known. Pressure and temperature at the surface of Mars is nearly always below the **Triple Point** of water. So it is rarely liquid there, although it may be at lower depths where conditions are suitable.

Ground Strength - Mars' surface generally consists of of a fine **Soil** containing sand and dust, with variable amounts larger rocks, and exposed bedrock (Figure 4.14-2). Surface strength in sandy areas is low enough to have trapped a rover, but in others is hard and sharp enough to have punctured metal wheels. Soil conditions come from a combination of impact cratering, volcanism, and more atmosphere and water early in the planet's history. As a result, suitability for construction and transport will be variable by site and need local investigation. The limits of overlying rock strength would be reached about 16 km in depth, but reduced by fracturing and water/ice fraction. Support structures would be needed for mining or drilling below these limits. Support structures are also needed for surface excavation which exceeds the local angle of repose for loose material.

Gravity Level - Mars surface gravity varies from 3.683 to 3.743 m/s^2, a 1.6% variation, with a reference value of 3.711, or 37.85% of Earth's. Lower values are due to equatorial location and extremely tall volcanoes, while higher values are at lower altitudes in the north polar region and Hellas Basin. Free fall conditions in orbit produce no net gravity, and gravity on the Moons is negligible. Structural support needed against gravity is therefore significantly lower than on Earth. Gravity levels needed for long-term human, plant, and animal health are unknown, but may be higher than Mars surface values and are definitely more than the levels in orbit. In both cases artificial gravity can be generated by rotation, in which case significant structural loads will be imposed on equipment.

Radiation Level - Radiation levels were measured by the Mars Science Laboratory/Curiosity Rover mission at 0.64 ± 0.12 mSv/day on the surface and 1.84 ± 0.30 mSv/day in orbit (**Hassler et. al., 2013**). These are 1.5 and 4.5 times higher than crew on the International Space Station get, and up to 200 times

Figure 4.14-2 - Surface of Gale Crater as seen by Curiosity Rover

the US annual average from all sources. These levels are unacceptably high for long-term occupation of the region. The simplest near-term solution is bulk shielding of a meter or more, using material from the surface or the moons. Long-term solutions include increasing atmospheric mass and an artificial magnetosphere.

Communication and Travel Times - Round-trip (ping) time to the Mars region from Earth varies from 6 to 45 minutes, depending on relative orbital position and need for a relay satellite to avoid the Sun. Travel times from Earth vary according to the relative positions of Mars and Earth, and the propulsion method chosen. When aligned, minimum energy transfer orbits typically take 8 months one way. As noted above, Mars's orbit has an eccentricity of 9.3%, so the maximum distance from the Sun is 20.6% more than the minimum. This makes reaching it from other regions a somewhat variable proposition. Orbit periods within the region range from 100 minutes at low altitude to 70 days at the region boundary. Minimum energy transfers between points will therefore take half these values. Actual travel times will depend on the methods used.

Stay Time - Average stay times affect transport and habitation needs. No one has traveled beyond the Moon yet, so there is no historical data for this parameter. Minimum energy trajectories between Earth and Mars impose transit + stay times of 1.5 to 3 years. However, these trajectories assume inefficient chemical rockets. Advances in technology and previous development of other regions of space should allow more options. So we will let future projects define this value based on internal needs until long=term occupation of the region establishes an average.

Transport Energy - Reaching the Mars region from Earth requires a theoretical velocity change of 11.8 km/s, and thus a theoretical kinetic energy of 69.6 MJ/kg. Actual energy required is generally much more due to inefficiency and overhead of the transport methods used. This is large compared to the 10-20 MJ/kg production energy of typical products. Therefore local production is favored where possible, and why we choose a bootstrap approach to developing Mars and other space regions. The planet's mass is 10.7% of

Earth's, and mean radius is 53.2% of ours. This results in an escape velocity from the surface of 5,027 m/s. Theoretical minimum energy to reach orbit is 5.5 MJ/kg and to the farthest point in the region is 12.5 MJ/kg. Efficient transport from the surface is therefore less energy than typical production energy, and it would be reasonable to supply materials to orbit from the surface.

Delivery in the other direction can exploit friction with the atmosphere, and therefore require less energy. Large-scale infrastructure like a space elevator can use regenerative energy, where the energy of cargo being lowered can be used to raise cargo in the other direction. With balanced traffic, in theory this requires no net energy. In practice traffic is not likely to be perfectly balanced, and real systems have overhead and inefficiencies. Transport between points on the surface, between points in the orbital region, and between the region and other destinations besides Earth will have highly variable energy requirements. They depend on the transport methods used, the starting and end points, and for orbital travel, on the departure and arrival times.

Available Resources

Our approach in this program phase, as in other phases, is to bootstrap development of the Mars region using a starter set of equipment brought from elsewhere. That equipment uses local resources, where possible, to make more equipment, and then finished products to use locally or export elsewhere. When local resources or production capacity are inadequate, needed inputs are supplied from other regions. The percentage of imports should decrease over time as the region develops, and exports increase. With both imports and exports, Mars would be economically integrated with the rest of civilization and may become self-supporting. To follow this approach, we first need to understand what local energy and material resources are available.

Energy Resources

Solar Energy - This is likely to be the main energy source for the region. Incoming solar flux at Mars varies from 494 - 716 W/m^2 (36 - 52.5% of Earth), with an average of 579.4 W/m^2. The range is due to Mars' variation in orbital distance from the Sun. The orbital region around Mars can differ in flux by ±0.33% from the planet, due to added or reduced solar distance. The distribution in wavelengths is the same as at Earth. Total solar energy in the region averages 213 million TW, or about ten million times civilization's 2017 energy use. This energy is available 100% of the time in orbit around Mars, except when crossing the planet's shadow. Low equatorial orbits can be shadowed nearly 50% of the time, while circular orbits at the edge of the region are shadowed at most 0.23% of the time, and much less if inclined to Mars' orbit plane. In terms of what can be done with this much energy, it could theoretically be used to completely dismantle Mars and turn it into other products in less than 2000 years. There would be immense practical challenges in doing so, but it shows the available energy should be enough for any future projects in the region.

Incoming energy to the planet averages 20,900 TW, or one thousand

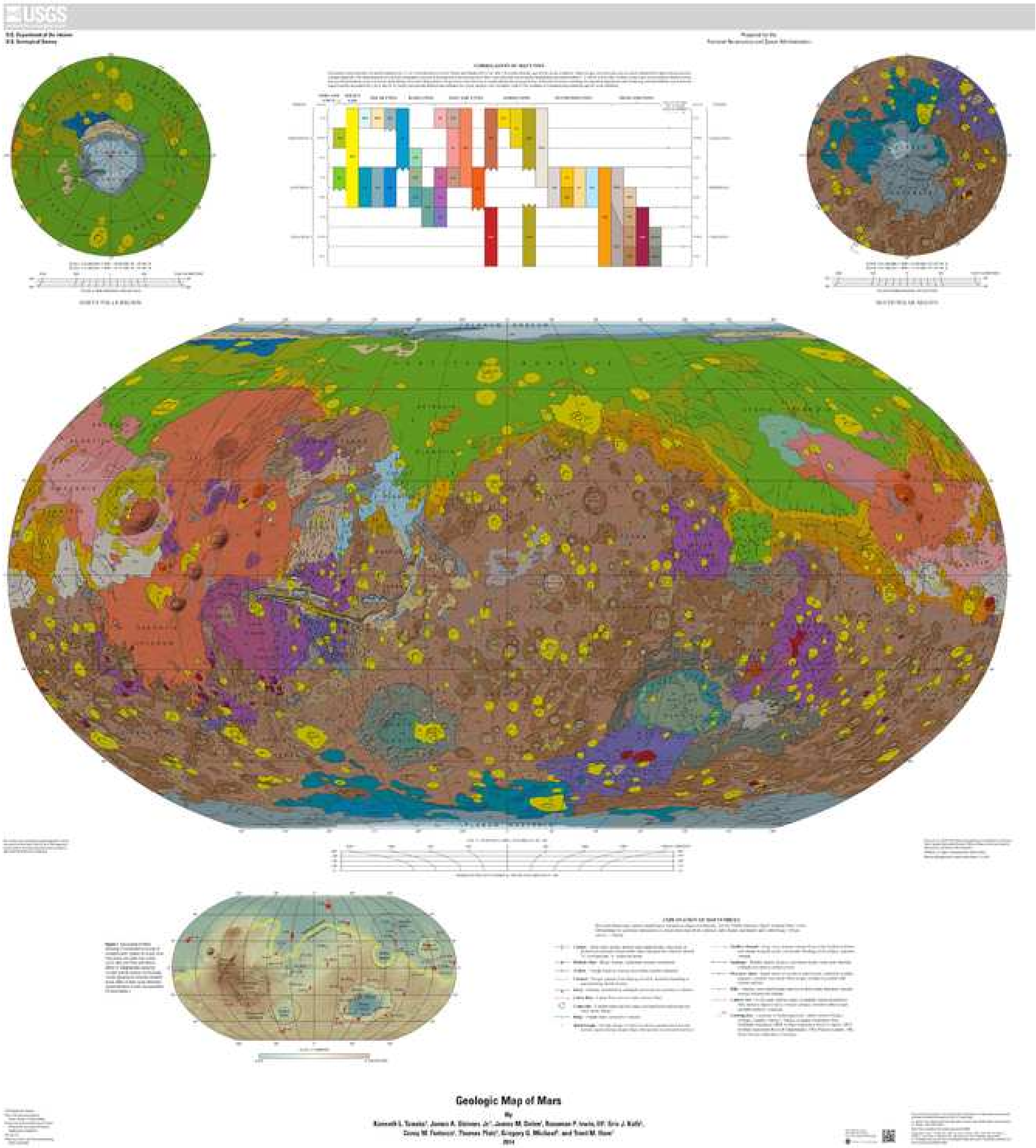


Figure 4.14-3 - Geologic map of Mars.

times Earth's 2017 energy use. Surface flux is reduced by a factor of 4 on average because a spherical planet has 4 times the surface area relative to the cross section which intercepts sunlight. The available amount varies by latitude, season, and time of day. It is further reduced by 20-40% at typical elevations from dust scattering and absorption in the atmosphere. Atmospheric gases have little effect because of their low density. Global **Dust Storms** average once per three Mars years, and regional ones annually. They can block as much as 99% of sunlight on time scales of a month.

Other Energy Sources - Orbital time in Mars' full shadow can be up to 3.8 hours, average nights on the surface are 12.3 hours, and locations above 65 degrees latitude experience long periods of seasonal darkness. Alternate energy solutions will likely be needed for for these times and during dust storms. Options include reducing operations temporarily, batteries, chemical energy storage of methane and oxygen made from local CO2 and water, thermal storage in local bulk rock, nuclear power, long distance transmission lines, and beamed power from orbit. Wind and geothermal are not likely to be useful, but will not be ruled out categorically until more analysis is done. Martian basalts may contain on the order of 5 ppm of Uranium and Thorium. If geologic processes have created concentrated deposits, they may be useful for local energy supply.

Material Resources

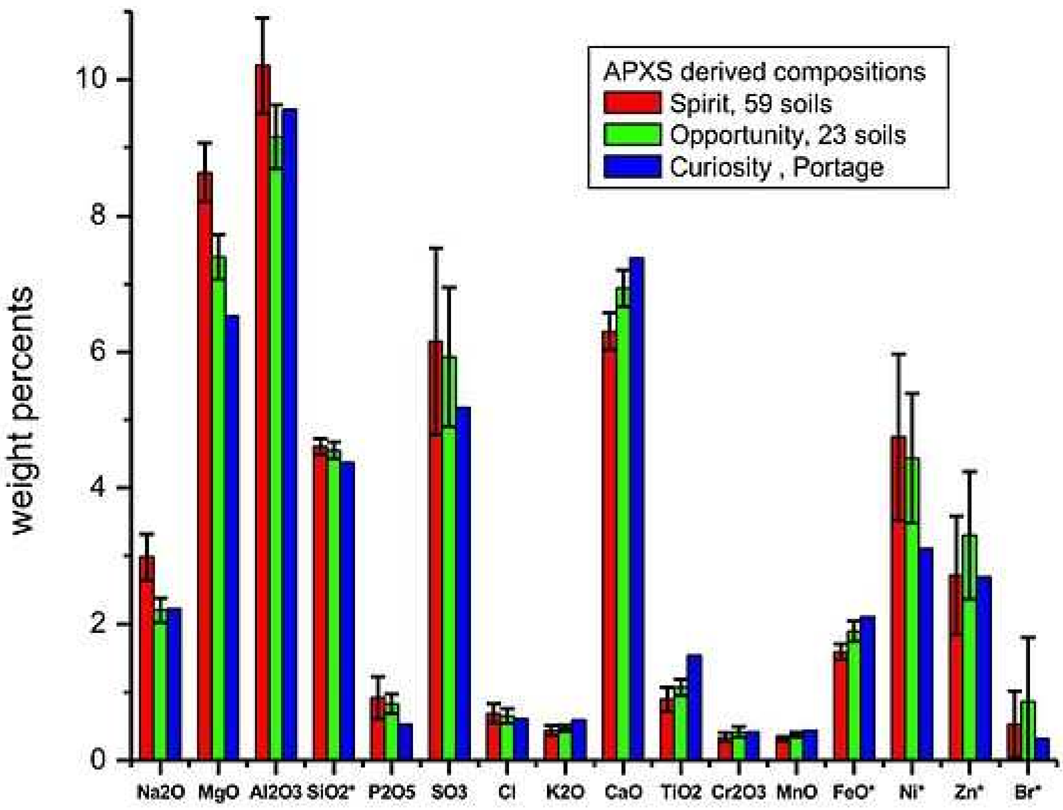


Figure 4.14-4 - Elemental oxide composition of typical soils at three landing regions on Mars. Error bars indicate the variation among samples. Note that for scaling the graph, the concentrations of silicon dioxide and iron oxide were divided by 10, and nickel, zinc and bromine were multiplied by 100.

Planetary Surface - Our detailed understanding of Mars is the result of a large number of orbital and surface instruments sent there since the mid-20th century, and over 100 **Martian Meteorites** delivered to Earth by natural processes. It is a differentiated terrestrial planet, with an iron-nickel-sulfur core, a less dense silicate mantle, and a relatively light crust. The surface has a **Varied Geology** as a result of internal melting, **Volcanism**, and other active processes earlier in the planet's history (Figure 4.14-3 and see **USGS Map 3292 (2014)** for a more detailed version). The surface composition, as measured by rover instruments, consists mainly of a variety of metal oxide minerals, with smaller amounts of volatile compounds like water and bromine (Figure 4.14-4). In terms of rock types, the original surface was primarily **Basalt**. This is made of different mineral crystals which form in a sequence as lava cools. Asteroids added their own components and mixed the crust by forming craters. Higher early levels of water and atmosphere caused oxidation, erosion, and chemical modification of the original materials. There is some evidence of **Tectonic Activity**, but much less than for Earth. These processes slowed after the first billion years as the planet cooled, the asteroid population was depleted, and water and atmosphere were lost to space.

The surface has large amounts of sand, dust, and smaller rocks making up the native soil layer (regolith). Early mining and excavation of the soil should be relatively easy. Hard rock mining and tunneling of outcrops and bedrock should be of comparable difficulty to Earth. Due to lesser gravity and thermal gradients, mining and drilling to depths of 30 km should be feasible with current technology, making over 4 billion km^3 of total material accessible. This is hundreds of millions times Earth's current annual mining needs.

Orbital Region - This is mostly empty aside from the two natural moons. Depending what materials are needed, they may have to be imported from the surface or other regions. Small amounts of solar wind particles and interplanetary dust pass through the region, along with occasional larger asteroids. Phobos and Deimos average 22.2 and 12.6 km in diameter, but are irregular in shape. Their escape velocities average only 11.4 and 5.5 m/s repsectively, so modified mining techniques will be needed to prevent loss of surface material and creating an orbital debris hazard. They have a combined mass of 12.8 trillion tons, or two centuries of Earth's total mining output, so represent a significant material resource in convenient orbits. Central pressure on Phobos, the larger moon, is only 120 kPa, or 1.2 times Earth atmospheric pressure. This presents no difficulty in tunneling or excavating to the center of the moon, so the entire volume of both moons are accessible.

The composition of the moons is uncertain based on current visible, and near and thermal infrared spectra. They resemble both those of outer-belt asteroids and some Mars surface minerals, but are not a match to either. One possibility is they formed from orbital debris after a large asteroid impact. The many large craters on Mars support this idea. The moons could later have directly accumulated asteroid material by impact, as their own collection of craters indicates. Resolving this question and determining their detailed composition requires closer observation.

The **Mars Trojans**, who share Mars' orbit, and tens of thousands of other **Inner Interplanetary Asteroids**, may be useful sources of raw materials. The latter group include the inner part of the Asteroid Belt, which Mars' orbit skims, and other asteroids orbiting somewhat closer to the Sun. Because of their large number, and the ability to use Mars for gravity assists, there should be some in particularly easy orbits to mine and return materials from. These asteroids can provide more varied source materials than the two moons alone. The planet can serve as an additional source of materials, but chemical rockets using a methane/oxygen mix would need several times the payload mass to reach orbit, which is not efficient. Alternate approaches for large-scale delivery from the surface include electromagnetic and centrifugal ground accelerators, and space elevators.

Industry Survey

Our next step is to identify possible economic activities which would justify development of the Mars region. Our approach is look at all existing Earth industry categories, and identifying ones that can potentially operate in the region at some point in time. To these we add any unique activities that only apply to Mars. The activities include those needed internally in the region, and products and services supplied to Earth or other space regions. We will use the latest version of the **North American Industry Classification System (NAICS)** for the list of categories, and apply their numbering system and sequence for our survey. Unique Mars activities are inserted as they are identified, under the most relevant heading.

11 - Agriculture: Wherever people live they need food, and closed ecologies can help recycle human wastes. Importing food from distant regions is likely to take more energy than cyclic agriculture, once it is set up. Biological growth can also supply other useful products besides food. So we consider agriculture a likely activity after the earliest stages of development. Sufficient sunlight is widely available in the orbital region, and a natural day-night cycle occurs on the surface. In orbit it requires filtering of harmful wavelengths, and in both orbit and on the surface, natural sunlight needs moderate concentration due to greater distance from the Sun. It is not obvious if artificial grow lights, such as **Light-Emitting Diodes**, can increase overall growth efficiency over natural light, but is worth investigating. Dust storms can block nearly all light at the surface, but food storage is well understood, and the natural temperatures on Mars are cold enough for refrigeration.

The major elements in typical plant tissue are oxygen (45%), carbon (44%), hydrogen (6.3%), nitrogen (1.3%), silicon (1.2%), and potassium (0.9%). The atmosphere, water, and soils on the surface can supply all of these. Of the remaining ~1% in needed trace elements, some are also available in the soil. All of the elements may need processing into forms that plants can use. Asteroids, and probably the Martian moons, have a wide variety of elements and compounds, so they can serve as source materials in orbit. Once efficient transport is developed from the surface, it can also be used as a supply. These materials will also likely need processing to usable form. Any remaining missing ingredients can be supplied from previous regions or Earth, but should be a small percentage.

21 - Mining: Extraction of local materials is a basic assumption of our program, because it takes less energy than importing everything from Earth. The Mars surface has large areas of broken rock (regolith), sand, and dust on the order of meters in thickness, which amounts to millions of tons/km^2. This is more than enough for early mining, although concentrations of specific useful ores is yet to be determined. The moons and nearby asteroids can supply enough materials for early use in orbit, and efficient transport from the surface can supply more materials later. We don't expect large-scale delivery of mined products to Earth, but there may be moderate amounts exported from Mars to nearby regions.

22 - Utilities: Nearly all activities in the region will need local power in some form. At first this can be supplied by imported solar panels and batteries, possibly supplemented by small fission reactors or thermal storage. Concentrating reflectors can supply direct heat and drive thermal cycle electrical generator. These can be built locally once manufacturing is developed. Solar energy is more available and more constant in orbit, so that may be the preferred location for energy-intensive activities. Beamed power from orbit to the surface may be useful as a supplement, especially at higher latitudes where sunlight is weaker and seasonally intermittent. Beam attenuation during dust storms is expected to be on the order of 50% at 32 GHz, and less at lower frequencies. So this may be a solution for lack of surface solar energy at these times.

23 - Construction: Relative to the Earth's land area, Mars' surface is nearly the same size, and the orbital region has 2,400 times the cross section. We use cross section rather than volume for orbit, because sunlight can only be intercepted once for power and natural lighting. Although the Earth will not be overcrowded from a purely physical space standpoint, environmental pressures and a desire for more personal room, flexibility in design, and novelty may drive local residential construction. Any non-residential operations involving people and other living things will also require suitable habitats, and some industrial activities will need protection from the local environment. All of these will in turn require some level of construction activity

31-33 - Manufacturing: Local manufacturing is also a basic assumption of our program. Any product which is easier to make in the Mars region than imported from elsewhere is a candidate, along with export products which are easier to supply from Mars than at their destination. At first only the most necessary, easy to produce, and high-leverage products would be made. These include items like radiation shielding, propellants, life support supplies, and basic construction and fabrication materials. Manufacturing would evolve over time, using bootstrapping, smart tools, and import of key items. Eventually all the local products which make economic sense will be made. Mars has significant amounts of Iron, Aluminum, Magnesium, Titanium, Chromium, Manganese, and Nickel for structural and machine alloys, Silicon for solar energy and electronics, and Sodium, Phosphorus, Sulfur, Chlorine, Potassium, and Calcium for chemicals. So it has a reasonable variety of material inputs for manufacturing.

42 - Wholesale Trade: We expect activities in the region will trade with each other and with other regions according to the principle of comparative advantage. We cannot predict at this time which specific activities in the region will have such advantages, but we can identify likely contributors. Transport from Earth or distant parts of other regions involves a lot of energy. Where local material supply and production can be performed for less energy, it would have an advantage. Surprisingly, it takes less energy to reach high Earth orbits from the Martian surface than from Earth. However, from a cost standpoint the break-even point is likely to be closer to Mars. More energy is available in orbit than on the surface, so operations that need high energy would prefer being there. Mining on land on Earth is relatively easier than in the oceans. The equivalent land area, and lower gravity and thermal gradient, makes several times the total volume of crust available to be mined on Mars. However we have little information about high grade ores as yet. Total available material is also about 4 times the mass of the Asteroid Belt, but compositions are different. So it is likely inter-regional trade will evolve based on relative scarcity.

44-45 - Retail Trade: We expect retail trade to take time to evolve on Mars. Early populations will be there for scientific and industrial reasons, and their personal needs for habitation space and life support will be supplied for them. As communities evolve past this early stage, more time and resources can be devoted to personal choice and non-essential items, and therefore suppliers at the retail level can evolve. It is hard to predict the form this will take, given the development of online marketplaces and automated delivery on Earth. It may be central warehouses rather than retail shop fronts. On the other hand, people may still want to select items in person, with the assistance of human staff.

48-49 - Transportation and Warehousing: Transport is necessary from other regions to start any activity on or around Mars, and is accounted for in previous phases from where such transport originates. Internal transport within the Mars region, and departing the region for other destinations belongs to this phase. Operation of transport fleets and infrastructure is covered under this activity, while their manufacturing and construction are covered under their respective headings. Storage and warehousing of all kinds is also included here. We expect all of these to continue in the region, so long as any kind of activity is present.

51 - Information: Delivering information through open space needs no mass, and little energy, so is relatively easy to do. Communications to and from Mars already exists through deep space networks to the spacecraft operating there. As remote control of equipment and human occupancy increase, local networks will likely be built. This can start with relay satellites in orbit to connect points on the surface, in the orbital region, and with other regions. Local networks will also be provided as various locations grow.

Local storage and processing will be needed for most modem activities. Equipment for these activities is complex and low mass, so at first will mostly come from Earth and older regions, with local production growing over time.

52 - Finance and Insurance: At first, most finance and insurance for the region would be handled from Earth. They involve non-material relationships between rights, contracts, and money which can be transacted remotely. Local offices and agencies may be needed for things like damage claim inspection. They would also make setting up new accounts and agreements easier due to the time delay communicating with distant regions.

53 - Real Estate, Rental, and Leasing: The Outer Space Treaty of 1967 prohibits national claims to celestial bodies, but ownership and use of private equipment, such as satellites, is already well developed. We expect the legal gray area about land, orbits, and mining rights in space to be settled in previous regions first. By the time significant development in the Mars region starts, the legal procedures needed should be available, but we cannot say at this time what they will be. Derived activities, like sale and leasing of satellites and their use, is already an active market around Earth, and such activities would likely be extended to Mars.

54 - Professional, Scientific, and Technical: Some activity in this category can only be carried out locally in the Mars region, and has already begun with scientific and technical research. We expect this to increase over time, and represent a large part of early Mars activity. To date, most of the people involved in these activities have been on Earth - astronauts in space are the rare exceptions. We expect this situation to continue in the near term, with gradually increasing numbers of people local to the Mars region over time.

55-56 - Management and Organizational Support: Business management and administration will likely be handled from Earth at first, for cost reasons. As the size and complexity of operations grow in the Mars region, time savings will start to favor local management. With increasing development of smart tools and networks, traditional organization trees may no longer be needed. Management and organizational support may be distributed and automated instead. So it is hard to predict the form this activity will take in the longer run.

61 - Education: The first people in the region will be educated elsewhere, and information systems are becoming advanced enough to support local training as-needed. Local education of young people would be deferred until permanent habitation exists with children being born and raised locally. The form it will take is unknown. By then it may be mostly via augmented reality devices rather than classrooms with human teachers.

62 - Health and Social Services: Due to the unusual conditions and hazards in the region, health monitoring and first aid capacity would be needed as soon as people are in the region. At first it would be by training the crews themselves, with remote monitoring and advice. As the local population grows, additional specialized equipment and staff would increase the level of care. Telepresence, artificial reality and intelligence, and haptic robots my be good enough by this time to provide health care remotely within the region, but signal delays are too large from other regions. Nursing and residential care would be provided on Earth at first, requiring return of people needing it. Once sufficient population is in the region, local facilities may be established. Social assistance would likely be handled remotely if needed. The basics like habitation and food need to be provided for everyone in the region by design, because the natural environment cannot support them.

71 - Arts, Entertainment, and Recreation: Entertainment can start with remote delivery and stored media for people in the region, because their mass and energy requirements are low. Activities like creative and dramatic arts would likely be deferred until surplus time and resources are available. Active recreation would begin with exercise for health maintenance, and may develop towards sports later. Early exploration and geographically unique locations may be preserved for future generations.

72 - Accommodations and Food: As noted under Health, basic living space, food, and drink are required for people in the region. Sponsoring organizations would provide them at first for crews, who would self-operate them. As local capacity grows and people establish long-term residences, there will be opportunity for rentals, temporary travel accommodations, tourism, and specialty food and drink. Tourism is a large industry on Earth, and the Mars region is unique enough to be attractive, but time and cost will limit voluntary travel at first.

81 - Other Services: This includes miscellaneous activities not covered elsewhere. Repair and maintenance will be highly desired for imported items. They are either expensive or slow to replace, and may be critical equipment. Once local production is established, surplus items can be stockpiled for replacement and repair needs. Personal services will start out self-provided. Private and civic organizations are likely not needed at first. Later they can be extensions of existing organizations, or self-organized locally.

92 - Public Administration: Support for this category will start from Earth or previous regions. Local fire and public safety would start out self-provided, and develop as specialties with larger populations. Environmental quality and monitoring are necessary functions and would be included by design. Publicly funded civil space activities are the only ones carried out in the region so far, and will likely continue to be important. National security activity has not been needed in the Mars region so far, and may stay minimal through agreements and collaboration. Private activity has not yet begun in the region, but once it does, public supervision will likely be necessary. For time and distance reasons, some of the supervision will be local. Public-private partnerships are quite possible.

Project Drivers

Many of the activities noted in the above survey will not begin until the middle or distant future, and technical and organizational changes are likely by then. So it is hard to predict which of them will make sense, and when. However, we can start to identify important factors which will drive projects in the region when their time comes. These include human motivations, economics, available technology, and the place of a project among prior, parallel, and later ones. We expect the importance and status of these factors to change over time, affecting which projects get started, and when.

Motivations

Motivations to develop the Mars region can be personal, organizational, or social. Human curiosity about the world around us drives current scientific exploration of Mars, and is likely to continue. Mars can eventually support other scientific work in other regions. The desire for safety from natural and artificial hazards is another major motivation for people. In this context Mars has been proposed as a backup location for civilization, in case something happens to Earth. Our program assumes much wider development of the Solar System, partly for this reason. Most of space is devoid of life and already full of radiation. So moving hazardous activities off our original planet to other regions would increase safety there.

In the economic realm, the desire for profits if ever-present. If they can be gained in the Mars region, that would be a strong motivation to develop it. Since the material and energy resources of the region are currently unclaimed, they would not need to be bought, merely exploited. [Additional motivations to be supplied].

Economics

Economics is a key driver for what projects will be started, their scope, and their timing. Currently, most activity around Mars is publicly funded. Governments and public institutions are limited by their available budgets, which pace what projects get initiated. Future activity at Mars is likely to have a large private component, which implies customers, markets, and the usual business considerations for what gets done and why. These include principles like comparative advantage and returns on investment. Private activities at Mars are more likely to happen if they are easier to do there than elsewhere, and can generate better returns on capital.

The natural advantages of Mars include a large and diverse source of materials, and the most Earthlike environment beyond the Earth's surface. Escape energy is 20% that from Earth, so efficient transportation is easier to build. By exploiting gravity assists and aerobraking, even transport to Earth orbit could take less energy than launch from the Earth's surface. Mars is distant compared to the earlier Earth orbit regions and the closest parts of the interplanetary region to us. However, it is close to the outer portion of the Inner Interplanetary region, from 1.25 to 1.8 AU, and would be a preferred source for that region. Since Earth already has a complete civilization, we don't expect Mars to supply large mounts of physical products. Trade would therefore depend on other types of value, or supplying nearby regions. An example would be using the Martian moons as a starting point for large-scale construction using low gravity and materials from both the surface and local asteroids.

Technology

Placement

Development Projects

The next step in our concept exploration is to combine the above information into a general approach for Mars development, and identify specific near and longer term projects to implement. We can provide early concepts for these projects, which will gives a sense of their scale and main features. However, this is merely a starting point, and does not exclude alternate ideas. It also does not include full optimization and integration of the projects to each other and to projects in other phases of the program. After describing our general approach, we list summaries of the projects by time, function, and location.. This is followed by a start at program integration. Where more concept details have been developed, they are provided as the last major portion of this section.

General Approach

Mars' surface area is nearly equal to all the land on Earth, and the orbital region has a cross section of 363 billion km^2, or 712 times the Earth's total surface. Like the Lunar region, the Mars region is therefore far too large to develop all at once, or by a single project or organization. Our general approach is then to identify a number of smaller tasks and projects. These can be put in a logical sequence, with later projects building on earlier ones, and be carried out by individual organizations or groups of them. These tasks and projects would interact with each other when they exist at the same time. They also interact with other program phases which are operating in parallel, and with the rest of civilization.

The various activities can be generally grouped by start time into preparation, orbital development, and surface development. Mars activities are less far along than those in Earth orbits or for the Moon, so most of the current and near-term work is preparatory, such as scientific exploration. We list these early activities after this general approach. Longer term projects follow them. Since they are not yet combined into an integrated sequence, we group them by primary function (production, habitation, transport, and services) and location (the orbital region or on the planet surface). Within each group they are placed in approximate time order. We expect many of these projects and activities to overlap in time, rather than be a strict sequence of one following another.

Preparation - Planning and designing future Mars projects requires understanding the features of the region in general, and of specific operating sites. Preparation for Mars development therefore involves tasks like exploration, surveys, prospecting, and site investigation. Scientific investigation of Mars, beyond merely determining its orbit, began as soon as large telescopes allowed seeing more than a point of light in the sky. It accelerated starting in 1964, when rocketry enabled sending instruments much closer to and landing on the planet, and communicating the data they collected back to Earth. We have not yet brought back pristine samples from the planet, but starting in 1983 over 100 Martian meteorites have been identified that came to Earth by natural means. These have been exposed to space and Earth environments, so are somewhat modified from their original condition. A number of spacecraft have orbited Mars, so the orbital region is now understood well enough to start development. Phobos and Deimos have not been orbited or landed on yet, and need closer inspection. A smaller number of landers and rovers have operated on the surface. Much more work is needed on the surface, since it is more diverse and variable than the orbital region. Exploring the whole planet will take a long time, so detailed investigation can at first start with proposed early landing sites.

Orbital Development - Orbital development is expected to lead work on the surface. This is because it is easier to reach from previous regions, and because of the existing material resources at Phobos and Deimos and nearby asteroids. Sites on the moons may need physical preparation, such as providing anchorage due to the low gravity or by gathering surface material to provide radiation shielding. Covers may need to be installed over active work areas to prevent loss of loose material. Open orbits elsewhere in the region don't need preparation, but also lack raw materials, so they must be imported. Due to long time delays from Earth, remote controlled operation would be difficult and slow. So we expect significant operations will require either advanced automation, or delivery of human crew to the region. These operations would begin with the delivery of starter equipment from previously developed regions, by means of electric tugs. An initial stock of supplies and propellants would also be delivered, but local production would be a priority in order to sustain operations. Phobos is likely to be used as a base of operations because of the large amount of materials available and convenient orbit to later reach the surface.

Surface Development - Once an orbital base of operations is built up, and a network of relay satellites are in place, site preparation on the surface can begin. At first this can use remote control of equipment from orbit. The time delays are short enough for near real-time control, and at first there will not be enough equipment on the ground to sustain people. A priority would be setting up a landing field and surface propellant production. Landers can then start shuttling from orbit carrying more equipment. When sufficient equipment is in place, crew visits can start, extending to full time occupancy. As in other locations, starter sets of production equipment are used to bootstrap larger scale production, then end-use products. Chemical rockets are inefficient, so in the longer term better transport infrastructure would be built up. Both orbital and surface operations would continue to get deliveries from previous regions, and start to export products once they have the capacity.

Current and Near-Term Projects

Research and Development -

Transport from Earth -

Transport from Low Orbit -

Missions to the Mars Region -

Long-Term Projects

As of 2017, a lot of R&D is needed before making definite plans for long-term development of Mars. Development of earlier regions will also significantly affect Mars projects because people, equipment, and supplies have to come from those regions. So the following list is only a summary of candidate projects, and, where known, what work is needed to prepare for them. As more work is done on Mars development, the list is very likely to change in terms of what projects are included, and their details. Linking these projects to each other and other parts of the program to form an overall plan is even more preliminary.

Mars Orbit Production

The Mars Orbit region is easier to reach than the surface, and requires low energy to reach from the Inner Interplanetary region which it is embedded in. We expect early production at Mars to start here, with basic products like radiation shielding, propellants, and crew consumables. Production build up would be an outgrowth of preparation stage tasks, such as remote control of exploration vehicles from orbit. Preparatory work would have delivered people, pre-made equipment, and supplies from prior regions to Mars orbit. Local extraction will reduce supply needs, and local production of equipment will further reduce external needs. Therefore it is highly desirable. The growth path would follow the standard bootstrap and self-expansion approach, using smart tools and local energy and material resources. As production expands, the orbital region can start to export to the surface and neighboring regions, both to support further development and generate economic returns.

Supply Sources - Large amounts of solar energy is available throughout the region, and is a logical way to power production. Availability is quite high except in lower orbits around the planet. The orbital region has two natural moons with a combined mass of 12.8 trillion tons. Using a combination of gravity assists, electric propulsion, and solar sailing, much of the Asteroid Belt and Inner Interplanetary region, with their thousands of asteroids, are accessible with low propellant use. Since the Earth Orbit and Lunar regions are previously developed, equipment and supplies can be imported from them. Eventually efficient transport from Mars would allow import of materials and equipment from the surface.

Processing - Aside from bulk radiation shielding, most products require conversion of raw materials by any of a vast selection of mechanical, thermal, chemical, electrical, or other methods. Equipment is needed to do these tasks and supply the energy to perform them. The growth path would then begin with processes that supply the most useful products relative to the amount and complexity of equipment needed. Equipment can either be imported or made locally. So the relative costs and difficulty of each source is a factor. Further growth pursues lower leverage processing, until it is no longer a benefit relative to product import, reaching an equilibrium with other regions.

Fabrication - Products like propellants and fluids don't require further fabrication after processing, just storage until used. However solid stock materials usually need additional work to make finished parts. There are a huge number of possible fabrication methods. Like the processing steps, they would be first be selected for highest output and simplicity relative to equipment and energy needs. Later expansion pursues lower leverage methods, until they are no longer an advantage relative to import.

Assembly and Construction - Finished parts, whether imported or made locally, are then usually assembled into finished equipment and facilities. Unless imported as complete units, the assembly and construction would be a local operation. Free fall in the orbital region, and very low gravity on the moons, allows large scale assembly with low effort. Some tasks may be easier to do under artificial gravity, so assembly and construction areas can provide it by rotation if needed. Due to relative ease of transport, large items may be built at dedicated factories, then towed to final operating locations.

Mars Surface Production

The Martian surface has a varied topography, climate, and geology. Therefore some places will be more suited to particular production activities than others. Early operations will not have the benefit of transport infrastructure, and will therefore have limited travel distances. Equatorial sites generally get more sunlight, and are easier to reach from the moons. These factors plus the choice of early production operations will guide site selection. Candidate sites would have been explored remotely with robotic equipment during the preparation stage. The first mining and construction equipment can then be delivered from orbit, and continuing remote operation until enough capacity to support people is in place. Early electrical power for production operations may be most easily done with stationary solar arrays, cables for stationary equipment, and charging points for mobile ones. Solar thermal energy can be used for operations that require heat. Additional energy sources can be added over time. Potential products include:

Minimally Processed Materials - This includes clearing and leveling for building sites and access roads by moving surface soils. Equipment which needs radiation, thermal, or debris protection would need excavation, placement, then covering with local materials. Unprocessed rocks can be used for slope retention and berms, and prefabricated arches can support soil coverage. Materials like metallic meteorites and useful ores may be gathered from the local area and stockpiled near a production site for later use. Bulk rock may also be used for thermal energy storage.

Water and Fluids - Water has numerous uses, including propellant production, life support, chemical processing, and in heat engines. Equatorial regions appear to have less water than the polar ones, but hydrated minerals may contain enough to be useful. It can be extracted by moderate heating and condensation. Permafrost ice can be extracted by heating soil above the triple point (0 C and 611 Pa), which is only slightly above ambient Mars daytime conditions. Electrolysis of water produces oxygen and hydrogen. The latter can be combined with carbon from the atmosphere to produce methane. This can be a propellant, fuel cell energy source, or feedstock for organic chemicals. Nitrogen can be extracted from the atmosphere and added to oxygen to provide a normal breathing mix.

Metals - **Steel** and **Cast Iron** are the most commonly used metals on Earth. They are primarily iron with 0-4% carbon, and other elements to produce **Alloy Steel** with different properties. The soil contains about 20% iron oxides, and the atmosphere is mainly carbon dioxide, which can supply the carbon. Manganese, chromium, nickel, and silicon are common alloying elements, and are present in Martian soils in significant percentages. Metallic meteorites are found on the surface and contain an iron-nickel-cobalt alloy which may be suitable as-is for early uses, but their quantity is limited. Large-scale production and more specialized alloys will need bulk reduction of mineral ores to provide the main elements. Magnesium, Aluminum, and Titanium are useful structural metals which are also present in reasonable percentages. Particular alloying elements for all the structural metals may need to be imported if they are not present locally in enough concentration.

Glass and Plastics - Common glass is made up of silicon, sodium and calcium oxides, and sodium carbonate. The first three are common components of native rock. Sodium is available, but mineral deposits of the carbonate form may be scarce, and need to be produced chemically. Pure silicon dioxide is used for quartz glass, which is useful for industrial processes. Native carbon compounds from which to make plastics are scarce on Mars, although carbon dioxide is abundant. Therefore chemical or biological processes would be used to make complex carbon compounds, or they would be imported from asteroid sources and the moons. Natural textiles would require large amounts of growing area, so are likely to be imported. Synthetic textiles are feasible with organic compounds.

Building Materials - Wood is unavailable on Mars unless it is imported, or the effort made to grow trees. Native materials may be used for clay-sand or sulfur-soil bricks and blocks. Concrete requires significant amounts of water and carbonates to form the cement binder. It is not clear if this will be a preferred process. Steel and basalt fiber can serve as reinforcement, although the latter needs an epoxy resin as a binder. For pure tension uses, bare fiber can be used, and it can also be used for fiber-reinforced metals. Mineral wool and vacuum-powder insulation can be produced locally.

Other Products - Electronics are low mass and high value, and are likely to be imported. Electrical goods like wires and motors need suitable metals, but are needed in enough quantity to make locally. It is not clear if coatings such as paints can be made from local materials.

Mars Orbit Habitation

Orbital habitats around Mars will share many similarities with those from previous regions, because the functions they perform and the operating environment are similar. Detail differences will result from things like lower solar flux and average temperatures, and a different mix of supply sources versus distance and difficulty. There would be a general trend from prefabricated to locally built and from small to large.

Prefabricated Habitats - We expect the first habitats will be delivered as completed elements from previous regions, along with an initial stock of supplies. They would either be delivered unoccupied by electric tugs, or carried along with a crew. Our general program approach is to use local materials and energy when possible. So we prefer an approach where the elements are delivered by the same kind of tugs used for asteroid mining, which is also the source for propellants and other supplies for the trip. We want the people to travel safely and efficiently, so we prefer a cycling transfer station approach. They get dropped off near Mars, then use high thrust propulsion for orbit capture. The habitat elements are already in position there, and are occupied in transit to a resource site, likely Phobos. An alternate approach is for the transfer station to grow by accretion of prefabricated and locally made elements. When it has grown sufficiently, part of the station separates, with enough crew and equipment to operate on its own. It then gradually adjusts its orbit to rendezvous with Mars. While it does so, it continues to mine local asteroids and get necessary supplies and equipment, which are delivered by separate tugs from Earth orbit.

Local Construction - Since some asteroids contain native iron alloy, an easy approach to local construction would be to heat that alloy in a solar furnace, then cast and machine structural shapes from them. Later construction can use more advanced materials and construction methods. Solar flux is about the same across the orbital region, so where to place the habitats would depend on the balance between access to asteroid, Martian moon, and surface materials. The surface requires a lot more energy to deliver from by chemical rocket, and has less energy supply than orbit, so at first the favored locations would be higher orbits. As orbital infrastructure is built up, closer locations using surface resources become more feasible. No single location will be best for all purposes, so we expect a number of habitats to be built. Larger habitats can be built by an incremental layering approach. The initial version would use several smaller prefabricated or locally made elements. A larger pressure shell is built attached to them. Once sufficient systems are installed, the larger shell is occupied. Later expansion adds additional larger shells around this core, which are occupied in succession as they reach completion. Later on, older shells can be dismantled and recycled to open up interior space.

Mars Surface Habitation

We expect surface operations on Mars to go through several stages. The one we are currently in is remote operation of equipment from Earth. Round-trip communications from Earth is 6 - 45 minutes, so this type of operation is necessarily slow. Once habitats occupied by people are in the orbital region, this will be reduced to less than 2 seconds, and from Phobos is 100 ms or less using a relay satellite to the far side. So near-real-time or real-time remote control will be possible. This will allow extensive preparation of surface sites, including preparation for people to be on the surface. Some activities will need people on the surface in person, and some people will want to be there for their own reasons. Once surface refueling capacity is available for landers, crews can visit the surface for short times, then return to orbit where more habitat capacity is available. After enough habitat elements are delivered and assembled on the surface, they can stay for extended times. Further construction will allow for larger resident populations. Various activities on the surface, like resource extraction, will likely need multiple operation sites. Therefore multiple habitats and surface transport between them will evolve.

The natural Mars environment is hostile to human life, so all living accommodations, from space suits to entire cities, must protect from the environment and one way or another provide for people's basic needs. These include air, water, food, temperature control, sleep, sanitation, and for longer stays radiation protection and artificial gravity. Small artificial environments, such as space suits and rover cabins, will have limited power and supplies. Therefore they will have time limits on occupancy, after which they will need to transfer occupants to a larger and more capable habitat, recharge, and restock.

Site Preparation - Habitation sites on the Mars surface are unlikely to be usable in their original state without some preparation. We expect that mining and construction robots will have been delivered for surface production, and remote-controlled from orbit. They can also be used to prepare the first habitat areas. Vehicle landings and launches, and propellant production and storage represent hazards. So the habitat sites should be located some distance away, and protected by local terrain, such as crater walls or hills, if available. Access roads will be needed between parts of the site. They would need clearing and leveling, and possibly gravel fill for load-bearing and traction.

Early Habitats - Landers carrying people will need at least a minimal crew cabin, but to save transport mass they would not be larger than needed. Crew can work out a lander cabin for very short surface stays, on the order of a few days. This is similar to how the **Apollo Lunar Module** functioned. Longer stays need additional supplies and equipment, which can be delivered ahead of time by uncrewed landers and unloaded by remote control. If needed, the habitat can go through a "construction shack" stage, where a minimal number of modules on a temporary site provide living quarters, while a larger and more permanent site is prepared.

A permanent habitat would likely be partly buried to provide radiation, thermal, and blast protection from transport or production accidents. It could take advantage of natural landforms such as craters, which are additionally cleared and leveled as needed. Both prefabricated pressurized modules and unpressurized vehicle and storage areas would be covered by arched supports, which are then covered by local soil and rock. For thermal and maintenance reasons, the modules would not rest directly on the ground, but on raised frames. The frames may rest on the ground if it is stable enough, or on foundation piles drilled and set deeper. At least one set of lifts and carriers will be needed to move and place the larger elements, and the final support frames should be adjustable to account for ground irregularities and settling. In order to provide pressure-tight living space, the various modules will need accurate alignment or flexible docking ports, but flexible sections are a weak point. It would be preferable to rigidly attach pressure shells. The habitat will need a variety of utility services, like power and fluid storage. These may be supplied from a production area if it is near enough.

Long Term Habitats - Zero gravity is a known health risk to people, but at present we don't have any data on partial gravity effects. We especially don't have any data on pregnancy and child development in low gravity, because no such astronauts have been to space. The 3/8ths of Earth natural Mars gravity may not be sufficient, even with added measures like exercise and body weights. In that case artificial gravity using rotating habitats on the surface would be needed. For ethical and child safety reasons, we should assume such facilities will be needed once Mars habitats reach the permanent colonist stage, until it is proven they are not required. For long-term stays, on the order of years, for adults who can make informed decisions, artificial gravity may not be needed. Artificial gravity is somewhat simpler in orbit, because it doesn't require mechanical moving parts, or dealing with friction and gyroscopic forces. So one possibility is restricting pregnant women and children from the surface, to avoid building rotating structures. But that imposes added transportation needs, and at present we don't have enough information to decide the best approach.

Mars Orbit Transport

Reusable Landers -

Orbital Tugs -

Mars Skyhooks -

Mars Surface Transport

Mars Orbit Services

Mars Surface Services

Program Integration

Concept Details

Phobos Base

Phobos has an orbit altitude of 5980 km (1.76 Mars radii) above the surface, and a mass of 10.66 trillion tons. This makes it a convenient starting point for orbital development, and an eventual base of operations for surface development. A Phobos base would eventually grow to include the full range of production, habitation, transport, and service functions. But it would start with the minimum viable scale and activities, and bootstrap up from these. Prior to starting, dedicated orbiter and lander missions should thoroughly survey the moon. Due to the low gravity, a lander may hop between multiple sites and get increased coverage. Problems with the **Philae** lander on a similar-sized comet may require a different approach, such as a "spiked ball", which can land in any orientation, then pivot to observe and use propulsion to hop to the next. Once the moon has been surveyed, we would know what set of starting equipment and supplies need to be brought from other regions. We expect that development of Lunar orbit and mining of the Earth's Moon and Near Earth Asteroids would have already begun. So initial deliveries would be via electric tugs from these regions. Additional raw materials may be imported from near-Mars asteroids if needed. Since Mars skims the inner edge of the Asteroid Belt, there are many nearby candidates.

Solar energy is available for at least 88% of the 7h 39m orbital period, with a 53 minute maximum eclipse. The moon is near-equatorial to to Mars, while the planet is tilted 25 degrees to its orbit. Therefore Phobos misses Mars' shadow entirely near the solstices, and has reduced eclipse times away from the equinox dates. The moon rotates once per orbit (i.e. keeps one face towards Mars), so surface locations have a day-night cycle. Energy-intensive operations may want to locate near, but not on the moon. This is to avoid loss of solar power from Phobos' shadow. Since the composition of this moon is not yet known we cannot plan out mining and processing in detail. At the least, bulk rock can be extracted for radiation shielding. Whatever the composition turns out to be, certain elements and minerals will be available, and we can find uses for whatever they are. Crater density and structure indicates a deep regolith, with a 1 meter layer of dust and small rocks, so excavation of this material should be fairly easy. If carbon or water are available in sufficient quantity, they can directly be used for propellants and life support. Otherwise, such materials would need to be imported.

A Phobos base would serve development of the orbital region and the surface in several ways. First, as a supply of bulk materials for shielding and counterweights for artificial gravity. These uses don't require processing. Next, whatever minerals and elements turn out to be there can be used in expanding local production and for end products. Some amount of materials would be imported from Earth and previous space regions, and later on from Mars, to make up whatever Phobos lacks. Over time, orbital infrastructure would be built up to enable larger scale and more efficient transport to and from the surface. The orbital vantage point allows for real-time control of ground equipment over the entire planet, with the assistance of several relay satellites. This is helpful before surface facilities can support human occupation, and later on as a relay network for dispersed ground locations. The resources of the Mars region can later support development in the Asteroid Belt and beyond.

Mars Skyhooks

Skyhooks are more efficient than chemical rockets for transport around large bodies. This is because they can use electric propulsion for orbit adjustment, which is about ten times more efficient, and store orbital energy for traffic going the other direction. In the limit of balanced traffic rates, they require no net energy or propellants to operate. They are easier to build than space elevators because they are physically smaller and lower stress, and therefore lower mass for any given task. However, they are still large-scale transportation infrastructure. Infrastructure is relatively expensive to build, but low cost to use each time. So they are not the first thing you build at Mars. Rather they are built when there is enough traffic to justify their construction.

We will look at three concepts for skyhooks as part of a Mars transport system. The first is is a pair of smaller skyhooks capable of reaching from low Mars orbit (LMO) to Phobos orbit, and provides whatever suborbital velocity to the Mars surface and velocity above Phobos that results. The second is a single larger skyhook capable of doing a full velocity transfer to the Martian surface. Third is an orbital skyhook combined with a ground accelerator. In reality, such systems can grow and evolve over time, and some other size or version may turn out to be optimal. It will take more detailed analysis and understanding what the traffic needs are before choosing an approach. For now, we just present these three as starting concepts.

LMO to Phobos Skyhooks

Orbit Mechanics - The product of a planet's mass M and the universal gravitational constant G is called the **Standard Gravitational Parameter** or μ. For Mars the value is μ = GM = 42.828 x 10^12 m^3/s^{-2}. This value is useful for calculating circular orbit velocities by the formula

$$v_o \approx \sqrt{\frac{GM}{r}}$$

The average radius of Phobos from the center of Mars is 9,377,000 meters (we must use all SI units and not multiples thereof), we can determine the orbit velocity of Phobos is 2137 m/s. Since Mars has an equatorial radius of 3,396 km, then Phobos is 5,981 km from the surface. For elliptical transfer orbits, where r is the current radius from the body center, and a is the **semi-major axis**, or half the long axis of the ellipse, the velocity at any point can be found from

$$v^2 = \mu \left(\frac{2}{r} - \frac{1}{a} \right)$$

We want our lower skyhook to avoid atmospheric drag, which starts to become significant at 160 km altitude. We then tentatively set our arrival at 240 km above the surface to account for the length of the lower skyhook. If we want to transfer from Phobos to 240 km above the Mars surface (r = 3636 km), then the velocities at the high and low points of the transfer orbit, and of the center of the lower skyhook can be calculate as follows:

- Transfer High point: r = 9,377,000 m ; a = half of high + low altitudes = 6,506,000 m ; from formula $v^2 =$ 2,551,844 m^2/s^2, and so v = 1597 m/s.
- Transfer Low point: r = 3,636,000 m ; a is the same as previous = 6,506,000 m ; therefore $v^2 =$ = 16,974,909 m^2/s^2 and v = 4120 m/s.
- 200 km skyhook center velocity: r = a = 3,636,000 m ; v = 3,451 m/s.

The velocity difference from Phobos to the transfer orbit is 540 m/s, and from the transfer orbit to the lower skyhook center is 669 m/s.These velocities are relatively small compared to the Lunar or Earth orbit Skyhooks we have looked at previously, so they would be relatively low mass ratio. Assuming the tips are at 1 gravity, the Phobos Skyhook would have a radius of 29.75 km, and the LMO one would be 45.6 km. Our initial assumption about an 80 km total length is then fairly close, and the structure only reaches below 160 km when vertical, so for now we ignore the difference. Phobos is tidally locked to Mars, always keeping one face to the planet, but it is not locked in rotation about the Mars-pointing axis, and tidal variations from the sun and slight orbit eccentricity cause it to wobble. Thus the Phobos Skyhook should probably not be attached to Phobos directly, but placed nearby

Both Skyhooks will have low mass ratios because of their low tip velocities. Therefore they would shift their own orbits by a large amount when transferring cargo. The solution is to anchor both of them with a sufficient amount of ballast mass from Phobos at their center points. The LMO Skyhook can drop cargo at it's own orbit velocity minus rotation velocity, or 3,451 - 669 = 2,782 m/s. Mars' equatorial rotation is 241 m/s, so the relative velocity to the surface will be 2,541 m/s. This is 73.6% of orbit velocity and 54% of orbit kinetic energy, which can be dissipated by drag. To reach Mars escape from Phobos requires adding 884 m/s. Since our Phobos Skyhook can add 540 m/s, that leaves 344 m/s to be done by other means. In total our two skyhooks can provide of the 2,418 of the 5,027 m/s from the Mars surface to escape, or 48%. This is a significant propulsive savings for such a small system, and mass ratios and payload fractions for the rocket portion of transport to and from Mars would be much improved.

Carbon fiber is an excellent material to build such structures. There are three possible sources for carbon to use in local production. The Martian moons, Phobos and Deimos may contain some carbon, but current observations indicate they are low in this element. No probe has yet visited these moons or observed them closely enough to be sure, so we cannot yet rule them out. Being in Mars orbit, they are a preferred source from a transport point of view. 75% of asteroids are the carbonaceous type, and there are many asteroids near Mars. The Martian atmosphere is 96% carbon dioxide. So those are alternate sources if the moons are not suitable. Basalt fiber is not as strong as carbon fiber, but the Martian surface is primarily basalt-covered. That is an alternate material to build with. Finally, for smaller skyhooks, transport from elsewhere with more developed industry is an option.

LMO to Surface Skyhook

A full orbit to ground Skyhook is a much larger system. It probably will not make economic sense until traffic grows to a higher level, but let us take a look at a possible design. Start by assuming a 1000 km high orbit. That will have a radius from the center of Mars of 4,396,000 m. From the above formula we calculate the orbit velocity is 3122 m/s. Subtracting 241 m/s for the rotation of Mars gives a relative velocity of 2881 m/s. If the tip is at 1 gravity centrifugal acceleration, then the Skyhook radius will be 846 km, and the tip will become motionless when the Skyhook is vertical 154 km above Mars mean surface level. That should be high enough to avoid significant atmosphere friction. The Mars Global Surveyor spacecraft used a no-drag holding orbit at around 175 km lowest point, and active aerobraking between 120 and 135 km. It did so when moving between circular and escape orbit velocities of 3,500 to 5,000 m/s. So the Skyhook with near zero velocity at the lowest point should not see much drag.

Structural loads vary from zero at the center to 1 gravity at the tips, and therefore average 0.5 gravity. Total stress on the structure is then 423 g-km. The highest strength known carbon fiber is 7 GPa, with a density of 1790 kg/m^3, and therefore a specific strength at 1 gravity of 399 g-km. Engineered systems are never designed at ultimate strength, because they fail at that point. Allowing a 2.4 design factor of safety, our design strength is then 166 g-km. Structural mass is exponential in the ratio of total stress to strength, so each arm of the skyhook is 11.78 times the payload mass, or 23.6 for the entire structure. This is a reasonable ratio for a system that will be used many times.

Escape velocity from circular orbit is 41% higher, and this skyhook can supply 93% at the upper point of rotation. Therefore it can capture payloads from, and inject payloads to, higher than escape velocity by a significant margin. However the chance of a missed capture raises a safety issue, since the payload would then fly past Mars. If it were carrying people, an actual design would either only capture from below escape velocity, or carry a backup method of slowing down.

LMO Skyhook with Ground Accelerator

Pavonis Mons is a 14 km tall shield volcano on the Martian equator. It has about 120 km of available slope in the direction Mars rotates. Assuming a limiting acceleration for people and cargo of 6 G's, a velocity of 3757 m/s is possible. This is higher than the 3313 m/s required for low Mars orbit, so either a shorter accelerator (93 km) or lower acceleration (4.5 G's) could be used. Atmospheric pressure at the top of the mountain is 140 Pa (0.14% of sea-level Earth), which is not an obstacle to acceleration, but will generate some drag, and a permanent orbit will need some propulsion to circularize. A ground accelerator this size would be a large construction project, so it makes sense to consider a combined system of smaller accelerator and a smaller skyhook instead of either providing the full velocity change to orbit.

Let us take the case of a low orbit skyhook capable of reaching Mars escape. It can then reach Phobos, Deimos, or other intermediate orbits by choosing where on the skyhook to let go. We start by guessing the orbit height is 360 km, to avoid the atmosphere and allow for the skyhook radius. Orbit velocity is found from the above formula to be 3377 m/s. Escape from that point is 4775 m/s, which requires adding 1398 m/s. At 1 gravity, the radius is then 199 km, or less than a quarter of the full orbit-to-ground version in the previous section. Total stress is 99.65 g-km, giving an arm mass of 0.86 times the payload, and a total structure of 1.73 times the payload. This is 13.5 times less than the full orbit case. Like other small skyhooks, it requires significant ballast mass to avoid throwing itself out of orbit when it accelerates a payload. The 199 km radius results in a 161 km altitude at the low point, and confirms our guess to avoid drag.

When we subtract the tip velocity from orbit velocity, which is what happens at the low point, we get a net velocity around Mars of 1979 m/s. Mars' rotation leaves 1738 m/s relative to the equator, or 49% of the 3554 m/s unassisted entry velocity. Therefore only 1/4 of the kinetic energy needs to be dissipated for re-entry, which should be easy. Going upwards, a ground accelerator at 6 G's has to be 25.2 km long, and longer at lower accelerations. This is 3.7 times shorter than the full speed accelerator, and requires the same amount less energy to operate. Thus both the skyhook and ground accelerator are significantly smaller in the combined case. The optimum division of work between the two, or whether to use only one or the other would depend on cost and design details, which it is too early to determine. We know enough at this point, however, to say a combined case should at least be considered for a future Mars transport system.

Depending on the required velocity, acceleration time, and cargo, a ground accelerator can be linear or centrifugal, and in the linear case can be electromagnetic or gas pressure-driven. The centrifugal case would be driven by an electric motor. Gas pressure and centrifugal approaches are more suited to bulk cargo at high accelerations, which makes the accelerator compact. People and delicate equipment are limited to 6 G's or less, which requires longer acceleration paths. In all three major concepts, the advanced transport systems can be built incrementally over time. Whatever portion is not handled by the advanced system is supplied by conventional rocket propulsion. Whatever portion is supplied by the advanced system will lower propellant needs and increase payload fraction relative to the all-rocket case. Those savings are the justification for building an advanced system. Skyhooks don't eliminate the need for propulsion. If their traffic is unbalanced, they still need some to maintain a stable orbit. But electric propulsion is at least ten times as efficient, so propellant needs are greatly reduced. If more traffic is going down than up, the skyhook will increase in altitude. In that case, atmospheric drag can be used to intentionally slow it down. Ballast mass can come from the Martian moons or nearby asteroids, and slowed down with aerobraking. It should therefore not require a lot of energy to put in place.

Mars Surface Systems

Construction

Earth-moving equipment will be needed for a number of purposes. The Mars surface is not protected from radiation like the Earth is, so long term habitats would need to be protected by a layer of soil. Landing areas will need to be flattened, and protective berms built around them so exhaust plumes don't sandblast nearby equipment. Once cargoes are delivered to the surface, they will need to be moved, lifted, and assembled, so devices to do those tasks will be needed. Most of the site preparation will likely be done by remote controlled machines.

Power Supply

Solar panels are a viable power supply on the Martian surface. It is rarely cloudy, aside from dust storms, and the atmosphere is thin, which partially compensates for the greater distance from the Sun. When larger amounts of power are needed, then radio-isotope or reactor devices can be added.

On-Site Propellant Manufacture

Producing propellant on the surface of Mars has been studied extensively, since it lowers the mass brought from Earth for a "Flags and Footprints" mission. If we already have a robust orbital mining and processing capability and Skyhooks in place to deliver cargo, there may not be much benefit in early production of fuel locally vs delivery. The economics of doing so will need to be examined. For portable power, such as in moving vehicles, and for rocket propellant to reach a Skyhook on a return mission, an Oxygen/Methane fuel mix is a reasonable combination. Once sufficient need for fuel exists, producing it locally will make more sense

Linear Accelerator

Pavonis Mons, which is located on the Martian Equator, has a slope about 175 km long, which rises about 6.5 km. If large amounts of cargo need to be delivered from Mars, a gas or electromagnetic accelerator can be used here. If the full slope is used, orbital velocity can be reached with human-tolerable accelerations (3.6 gravities). This would not be an early system, since sufficient traffic is needed to justify such a large installation. Another option is a centrifugal catapult on top of the mountain for early cargo launch.

It is quite feasible to build a rotating space elevator (Rotovator) in orbit, coupled to a linear accelerator on Pavonis Mons (http://upload.wikimedia.org/wikipedia/commons/1/12/Pavonis_mons_topo.jpg). You have 60-120 km of ramp space, and no atmosphere to speak of, so at 3 g's and 60 km you can reach half of Mars orbit velocity, and the Rotovator provide the rest.

Long Term Development

Often the phrase "Terraforming Mars" has been used in the past. This is not a good phrase because it means "Make Mars like Earth". Because of orbit and mass differences, we cannot make Mars just like Earth, nor do I think that should be the goal. I prefer the word "humanize", meaning making it more suitable for humans. It may also mean modifying humans to better suit the Mars environment (like the lower gravity). Large scale changes to Mars should be delayed till after we have a firm idea if there is any native life on the planet, and even then done with due consideration and forethought. They should also be delayed until there are enough people on Mars to justify the large-scale projects. So what follows is more to answer what is possible from a technical point of view, and less to say "I urge you to do all these".

Magnetosphere

Mars lacks a strong magnetosphere - a magnetic field around the planet that traps and diverts charged particles from space. The Earth has one due to the magnetic field generated by our planet's core. A strong magnetosphere protects the atmosphere from being slowly stripped off as solar wind and other particles hit the upper atmosphere. Short of stirring up the planet's core, there may be some other ways to generate a field. The practicality of any of them is yet to be determined:

- Run one or more superconducting cables around lines of latitude, which, like any current-carrying wires, will generate a field.
- Place some number of iron-nickel asteroids in Mars orbit and magnetize them, and point their fields in the same direction.
- Mars is red because there is a lot of iron oxide on its surface. Extract the iron, and magnetize it. You might be able to use the iron for other purposes at the same time as it being a magnet.

Magnets to make the magnetic field have fewer ways to break than superconductors, but if the superconductors work 99% of the time the other 1% doesn't make much difference to long-term atmosphere loss. Some leakage of the atmosphere will still happen because Mars is a smaller planet than the Earth, so it is easier for atoms to escape.

Greenhouses

If you want to eliminate leakage, and bring up the pressure to breathable levels without importing a planet's worth of atmosphere, you can use greenhouse domes. If you really need the space, you can extend the domes to cover the entire planet bit by bit. To create Earth sea level pressure on Mars, a pressure balanced dome would consist of 10 meter thick quartz, glass, or equivalent, which you extract from Mars surface material. Lighter domes tend to float up, as the internal pressure is higher than the surrounding air. In that case they need to be tied down so they don't float away. A very large or planetary dome doesn't need much to hold it up, just some towers or cables to keep it from moving sideways.

You can design the clear material like armored glass to be resist damage, and ten meters of anything is pretty hard to break. But anything can be broken, so a lot of thought needs to go into how to deal with damage. As a greenhouse, you can take advantage of the "greenhouse effect", which is the trapping of infrared heat radiated back from the ground. You can specifically select the glass type or add coatings to trap infrared. You also want to block Solar UV radiation, which is not blocked by the Martian atmosphere. On Earth the greenhouse effect is a problem, since we don't want the planet warmer than it already is. On Mars it's a solution, since it's too cold for us there at the moment.

Full Atmosphere

If you find living under a dome objectionable, you would need to provide a full atmosphere at a breathable level. That is a very big job because planets are large. On Mars you need to provide 25 tons of atmosphere for each and every square meter of the planet, or 3.6 million billion tons total. That's to provide one Earth atmosphere pressure. If you are satisfied with less oxygen (similar to mountains on Earth), and a different mix for the rest of the air, you can get by with somewhat less. Despite it's distance, the easiest place to get enough nitrogen might be the Kuiper belt, which is outside Neptune's orbit and which Pluto is a part of. You could use a "reverse gravity assist" from Neptune to drop the material into the inner solar system. Nitrogen is rather scarce in the inner solar system, and getting it from anyplace with a deep gravity well (like Earth) takes a lot of work. Some of the outer moons might have enough ammonia (NH_3).

[Text to be merged]

The combined systems discussed in the previous sections are a different route to getting to Mars. They build capabilities step by step, each one preparing for the next, and generally using machines to prepare the way for humans. In this section we will discuss the last steps to get to Mars, building on the prior technology from our combined system. Mars is the most nearly Earth-like planet we know of, so we will also mention some ideas for long term development. They would get done, if ever, much later, when Mars and the Solar System are more fully developed.

Section 4.11: Phase 5 - Planetary System Development (continued)

Phase 5 in general covers the larger Solar System bodies besides Earth, the moons and rings which many have, and the orbital regions where their gravity dominates over the Sun's. The large gravity wells around these planets require extra transport systems and energy to navigate. Sunlight is blocked sometimes in closer orbits, at night on surfaces, and always below the surface. Some of the planets and moons have atmospheres and trapped radiation belts, and all the larger ones have significant surface gravity. All of these conditions are different enough from the orbital regions in Phase 4 to need design changes. So we assign development of the larger bodies to a new major phase. The various planets and their surrounding regions are also different from each other, so Phase 5 is divided into five parts. They are in approximate order of difficulty and start times, and tend to start after the respective orbital regions where each body is located. They operate in parallel with all the other phases once started.

The first two parts, Phase 5A: Lunar Development and Phase 5B: Mars Development, cover the easiest to reach bodies from Earth. Concepts for these regions are developed enough we devoted two previous sections of this book to them - 4.9 and 4.10. Concepts for the remaining three are less developed and presented below. Phase 5C covers the hot planets Venus and Mercury and their surroundings. 5D covers Jupiter and the varied bodies and difficult environment around it. Lastly, Phase 5E treats the outer three Gas Giants, and the objects and cold environments around them. The Mars region requires slightly more velocity to reach than Venus, but Venus has a much deeper gravity well and no moons. So it is harder to reach anything useful at Venus, and conditions are much worse, so we place it and Mercury after Mars in our sequence. The Moon and all the major planets in Phase 5 have been visited by spacecraft, sometimes many of them, so we have reasonably detailed knowledge about them.

Phase 5C - Venus and Mercury Development

This phase includes development of the two major planets which are closer to the Sun than Earth, and orbits around them with semi-major axes less than 600,000 km for Venus and 100,000 km for Mercury. These are distances to which reasonably stable orbits are possible against the Sun's influence. **Venus** and **Mercury** have both been seen since ancient times, although it was not until the **Copernican Revolution** that they were understood to orbit the Sun. They are more difficult to reach than Mars. Added to this are the high temperatures for both, and high pressures for Venus. They combine to place this phase after 5B Mars Development in start time.It also comes after Phase 4C Inner Interplanetary Development, since you must travel through that region to reach the inner planets.

The similarity in size and orbit to Earth made Venus popular in **Fiction** until the literally hellish temperatures and lack of water were confirmed in the 1960's. Primary interest then shifted to Mars as the next most earth-like planet. Mercury has also received some**Attention in Fiction**, but less popular interest because it was known to be too hot and too small to be like Earth. In contrast, our program considers all of the Solar System's esources to be useful, limited only by the difficulty in accessing them.

Region Features

Venus is the brightest planet as seen from Earth. It's orbit is nearly circular at 0.723 ± 0.005 AU, coming closer to Earth than any other major planet. Since 1961, a total of **43 Spacecraft Missions** have attempted to reach Venus in some way, with 27 considered at least partly successful. It is 18.5% lower in mass and 5% smaller in radius, making it a near-cousin of Earth in overall size. However the atmosphere has a reference pressure 90.8 times higher than Earth, of which 96.5% is carbon dioxide, 3.5% nitrogen, and some trace gases. Combined with 90% higher solar flux, this results in an extreme **Greenhouse Effect**, raising surface temperature to 737 K (462 C). This is nearly hot enough to visibly glow red hot if you could survive the surface conditions to see it. The black-body temperature in sunlit orbits is 468 K (195 C), 75 degrees warmer than for Earth. Objects of other colors are similarly warmer. Surface gravity is 8.87 m/s^2, or 9.55% lower than Earth. While surface conditions are hostile, moderate pressures of 0.5 times Earth and temperatures around 27 C exist at 55 km altitude. These can be reached by tall or floating structures.

Mercury orbits closer to the Sun and is therefore harder to reach from Earth. Only **Two Spacecraft Missions** have visited it so far, with a third planned to launch in late 2018. It is only 5.5% of Earth's mass and 38% in radius, making it the smallest inner planet. The orbit is significantly eccentric, varying from 0.307 to 0.467 AU from the Sun. The axis is very nearly perpendicular to the orbit plane, so some polar craters can have permanently shadowed areas. The surprising result is despite being close to the Sun, these parts of Mercury can be as cold as 100K (-173 C). Conversely, the subsolar point at perihelion can reach 700 K. Since Mercury has nearly no atmosphere, heat is more easily radiated to space. So the highest temperatures are slightly less than Venus, despite being closer to the Sun. Sunlit orbital temperatures range from 575 to 710 K (302-437 C) for black bodies, and are lower for other color objects. Mercury's surface gravity is 3.7 m/s^2, the same as Mars despite being considerably smaller. This is because it is 33% denser, making it the second densest planet after Earth.

Direct transfer orbits from Earth to Venus and Mercury need about 2.5 and 7.5 km/s insertion velocities. These can be lowered by gravity assists. Escape velocities are 10.36 and 4.25 km/s from their surfaces and 1.041 and 0.664 km/s from the outer edge of their regions. Orbital velocities are 29.3% lower in each case. Orbit velocity from the surface of Venus is a theoretical number, since the thick atmosphere would impede direct launch. Orbit periods in the region range from 90 minutes to 59 days around Venus, and 85 minutes to 15.5 days around Mercury, depending on orbit size. Travel times from Earth by direct transfer orbits are 4.8 and 3.5 months respectively. Gravity assisted trajectories need less propulsion, but take longer. Round-trip communications times from Earth vary from 4.3 to 30 minutes for Venus, and 9 to 25 minutes for Mercury. Communication times within their regions are less than 8 and 1.3 seconds respectively.

Both Venus and Mercury rotate slowly, taking -243.02 and 58.65 days with respect to the stars. Venus is negative because it rotates opposite the orbit direction, unlike Earth. Their solar days are 116.75 and 176 Earth days, respectively. The low equatorial rotation velocity makes a negligible difference in taking off or landing on these planets. The lack of atmosphere and long solar night on Mercury produces large temperature swings. Venus' thick atmosphere has impeded some observations, but topography has been mapped by radar from orbit. We know less about global surface composition, but landed instruments indicate at least two types of basalt are present, and radar mapping indicates extensive **Volcanic Activity**. **Mercury's Geology** is influenced by a large iron core, volcanism, and extensive impact features. The surface has a primary composition of 40% Oxygen, 25% Silicon, 11% Magnesium, 6% Aluminum, 4% each Calcium and Iron, and 2% Sulfur, with some variability by location. This makes it more similar to the Earth's mantle than our crust.

Development Projects

The inner planets present some difficulties in reaching them, and from local conditions, but they have an advantage in large amounts of energy for production and other activities. This comes from 1.9 times Earth's solar flux at Venus, and 4.6-10.6 times at Mercury. To see how much an advantage this is, it takes 7 to 12 km/s velocity change to travel from near Earth to the Venus and Mercury orbital regions. High efficiency electric propulsion will use 0.15-0.27 kg of propellant/kg cargo without gravity assists, using 263-474 MJ/kg of solar array energy. If the cargo is also solar arrays, or equivalent thermal power generation, they will produce an additional 13.75 and 91.5 MJ/kg/day in added energy output, and repay the extra energy to get them there in 19 and 5.2 days respectively. Space system lifetimes are nominally 15 years, so energy-intensive processes highly favor going closer to the Sun. When added to the material resources of these planets, we have enough reason to consider development.

Orbital Development

Near-term development of Venus and Mercury would begin in orbit by delivery of ready equipment and supplies from previous phases, and bootstrapping further growth using starter sets of production equipment. Solar energy is available 60-100% of the time in orbit, and intensity is high, so it should enable rapid growth. Materials would initially come from asteroids in the Inner Interplanetary region, then supplemented with mining Venus' atmosphere and the polar regions of Mercury. Long-term development includes rotovator-type spaceports to make surface access easier, and building sunshields to terraform the planets below.

Orbital Production - Since Venus and Mercury appear to have no natural moons, we would like to use nearby asteroids as a resource for orbital development. There are a relatively small number of known asteroids near the orbits of those planets (the two inner circles on a **Plot** by the Minor Planet Center). It is not clear if this is from low scattering efficiency and short residence time for asteroids starting from more distant orbits, or the difficulty in spotting small objects in the Sun's direction when looking at their partly unlit sides. If nearby asteroids are insufficient, or the wrong composition, materials can be imported from better-supplied interplanetary areas with the help of planetary gravity assists. Venus' atmosphere is a ready source of carbon, oxygen, and nitrogen, assuming scoop-mining from orbit is developed. Mercury is small enough to directly throw bulk material to orbit, and the polar regions have tolerable natural temperatures. Equipment that needs to operate at Earth-like temperatures can be protected by sun-shields when local conditions are too hot.

Orbital Habitats - Habitats with artificial gravity and sufficient thermal and radiation shielding can be built in previous regions which are more developed, and transported to orbit around Venus and Mercury. This transport can be gradual, with modifications made as the conditions change, and supplies extracted as needed en-route. This approach can enable substantial occupancy from the start. Most of their design should be unchanged from previous orbital phases.

Orbital Transport - Orbital transport requirements include destinations within each orbital region, access of the planets below, and trade between the two planets and with elsewhere. The early needs can be met by extending Inner Interplanetary transport that was already operating. Later systems would be based in the local orbital regions. The previous systems would include solar-electric propulsion and planetary gravity assists. Solid materials are easier to extract from Mercury, and atmospheric gases from Venus. So trade is likely between those planets.

Increased gravity as you get closer to the Sun implies more energy is needed to change orbits and reach Venus and Mercury. Solar flux increases faster then velocity changes, and this introduces the possibility of solar sails as effective transport method in the inner regions. For example, reflected sunlight provides 15.5 Newtons/km^2 at Venus, and a 1 micron Magnesium-Aluminum sail material would mass 2400 kg/km^2. This generates 558 m/s/day acceleration for the bare sail. This is reduced by the remaining structure and cargo mass, and angling the sail to control thrust direction. This acceleration is comparable to that for electric propulsion including solar array mass near Earth. The advantage of solar sails is they do not consume propellant. Their disadvantage in the outer Solar System is low solar intensity makes them very slow. Electric propulsion is also quite viable closer to the Sun, and a combined system is possible to take advantage of the reduced propellant from the sail and wider thrust angles from the electric engine. In the long term, rotovator-type orbiting spaceports can assist with reaching the surface or escaping from Mercury and Venus, once enough traffic exists to justify their construction.

Orbital Services - [TBD]

Surface Development

Near-term development of the inner planet's surfaces is impeded by their generally hostile conditions. The exceptions are the polar areas of Mercury, and high altitudes in Venus' atmosphere, where conditions are more moderate. For full development, some level of terraforming is desirable, as discussed below. Higher latitudes on Mercury get less sunlight per area as a function of Sun angle, and terrain features or artificial reflectors can provide protected areas that are cooler. Either floating structures, or ones supported by towers, can reach moderate conditions at higher altitudes over Venus. It is not obvious which will be more practical, or if surface development on Venus should wait until some level of terraforming is accomplished. In the long term, if the surface conditions can be made more tolerable, primarily with sunshades, then large scale access to the surface would encourage development there. The combination of larger sources of raw materials plus continuous high energy available from orbit would make them attractive locations, at least for industry.

Concept Details

Terraforming Venus and Mercury

Venus and Mercury are mostly too hot, and Venus has too much atmospheric pressure, to use their surfaces in their natural state. A long-term project would be to modify these conditions to make them more Earth-like, a process known as **Terraforming**. **Terraforming Venus** has been considered by various means. Our approach assumes that asteroid and Lunar mining in previous phases becomes well developed. Material extracted from these sources is turned into many orbiting sunshades, which block most or all of the sunlight reaching the planets now. The shades would be in medium orbits around the planet and be facing the Sun while crossing in front of it. The remainder of their orbit they orient themselves so as to counteract the light pressure which would otherwise push them out of the desired orbit. To do this effectively they may need to be reflective on at least one side. Early sunshades can simply block the Sun, but later ones can incorporate solar collectors and other equipment to make use of the available energy.

The minimum mass to totally block the Sun can be estimated by the cross section of Venus and assuming 1 micron thick shades. This would only require 0.115 km^3 of material. Due to the shades being in orbit more than this is needed in practice, so we can use 1 km^3 as a more reasonable estimate for that planet. The largest metallic asteroid in the Main belt has 6 million times this volume, so there is more than enough material available for such a project. Mercury has only 16% of Venus' cross section, and needs correspondingly less reflector area, but the solar intensity is much higher trying to dislodge them. So it is not clear what the relative difficulty will be. Since Mercury has no atmosphere, sunshades installed on the surface may be a better approach.

Blocking the Sun for Venus would allow the atmosphere and surface to cool. A cooler atmosphere has a shorter **Scale Height**, over which pressure changes by a factor of *e*. Therefore high altitude terrain, like **Ishtar Terra** will preferentially see lower pressures and temperatures. Construction of towers on these high points would additionally lower pressure and temperature, making them the earliest places to occupy, with expansion to other areas as conditions improve. Some minerals, like **Peridotite**, can capture carbon dioxide, which makes up 96.5% of Venus' atmosphere. If such minerals exist on or near the surface of the planet, natural or accelerated capture may lower pressures further. Landers on Venus have detected basaltic-type surface compositions, so the right types of minerals may be present. Much further work on the **Geology of Venus** is needed before the feasibility of carbon capture is determined.

At first, more shading or entirely blocking the Sun would accelerate cooling the planet. Afterwards, the level of shade can be adjusted to maintain desirable temperatures. Mining the atmosphere from orbit using scoops can begin long before the planet cools. Since the atmosphere is nearly 0.01% of the entire mass of Venus, this mining it is unlikely to make a significant difference in the terraforming process. Even so, the carbon, oxygen, and nitrogen extracted this way have lots of uses. The shades can start out as simple reflectors, but as the orbital region develops, they can be gradually replaced by solar collectors and habitats.

Phase 5D - Jupiter System Development

Jupiter is one of the wanderers of the night sky known from ancient times. The discovery in 1610 of four moons going around it, rather than Earth, helped establish the modern **Heliocentric** (Sun-centered) model of the Solar System. Today we recognize Jupiter as the most massive planet in the Solar System, 317.8 times Earth, with 0.066 more in the moons and ring system. This is 2.5 times the mass of all the other known planets combined. Effectively it is a miniature solar system of its own. Included in the Phase 5D region is the planet, 69 **Known Moons**, a thin **Ring System** among the four innermost moons, and orbits with semi-major axes within 20 million km of Jupiter. The outer irregular satellites orbit farther than this, but we also include them as part of the Jupiter System.

This phase follows 5C: Venus and Mercury because it is more difficult in several ways. Jupiter orbits 5.2 times farther from the Sun than Earth. With the addition of navigating its gravity well, more total propulsion is needed to reach its resources. At the same time, solar flux is much weaker, and parts of the region have high radiation levels. The phase also follows 4D: Main Belt and Trojan, because Jupiter is at the outer edge of that region and sits between the Trojan clusters which lead and follow it. So it is a small step from the Trojans to developing the outermost Jupiter moons. For previous planets like Mars we divided development into orbital and surface projects. For Jupiter, we divide it into three parts. The Outer System includes the irregular satellites and orbits larger than 2 million km. The Inner System includes the four large moons, four smaller inner ones, the rings, and orbits closer than 2 million km. Finally comes Jupiter itself, but that development would be far in the future due to extreme difficulty in reaching it.

Region Features

Jupiter

Jupiter's orbit ranges from 4.95 to 5.46 AU from the Sun, with a period of 11.86 years. Travel from Earth by direct transfer orbit nominally takes 2.73 years. It has been observed by astronomers since the invention of the telescope, and **Visited or Orbited** by 9 spacecraft, including one currently active, and one atmospheric probe. It is a **Gas Giant** with approximately the same overall composition as the Sun, and therefore does not have a well-defined surface. The radius to the visible clouds is 71,492 km at the equator and 69,911 at the poles, giving it a total surface area 121.9 times the Earth. The difference in radius is due to the rapid rotation period of 9.925 hours, and that it is mostly not a solid body. The **Upper Atmosphere** is about 90% hydrogen, 10% helium, 0.3% methane, and small amounts of ammonia and other trace gases. Temperatures are 112 K (-161 C) at the 10 kPa level and 165 K (-108 C) at the 100 kPa level (about 1 Earth atmosphere). Conditions reach 2.5 MPa pressure and 430 K (157 C) temperature at a depth of 132 km below the 100 kPa level, which is where the Galileo probe stopped transmitting. Therefore there is a level where temperature and pressure are not too different from Earth, but getting and surviving there would be very difficult.

Solar flux at Jupiter varies from 4.1 to 3.35% of that at 1 AU, with an additional 0.2% variation across the orbital region. Black body temperature in the region averages about 173 K (-100 C). Escape velocity from low orbit is 59.5 km/s and orbit is 42 km/s. Since the equator rotates at 12.6 km/s, the required velocity to reach orbit or enter the atmosphere is 29.4 km/s. This large difference between orbit and the atmosphere makes accessing the planet itself very difficult. Round-trip communications time from Earth varies from 1.06 to 1.85 hours.

Inner System

The inner system extends from the lowest stable orbits to ones with semi-major axes of 2 million km. Orbit periods range from 2.97 hours to 18.28 days. It includes the four large **Galiean Moons** (Io, Europa, Ganymede, and Callisto), which orbit 0.422, 0.671, 1.07, and 1.88 million km from Jupiter in nearly circular orbits. The first three have 4:2:1 resonant orbit periods. Four smaller moons have orbits between 127,000 and 222,000 km in radius, with a thin ring system among them. Their depth in Jupiter's gravity well makes access difficult.

The large moons range from about half to twice Earth's Moon in mass. This is enough to be useful for gravity assists, making travel within the Jupiter System easier. They are 90-150% of the Moon's diameter, with a combined surface area of 232.8 million km^2, or 1.56 times the land area of Earth. All four are tidally locked to Jupiter, so their days are equal to their orbit periods of 1.77, 3.55, 7.15, and 16.7 Earth days. To reach them unassisted from beyond the Jupiter System takes 3.4-7.15 km/s of velocity change. Their escape velocities range from 2.0-2.75 km/s, and surface gravity from 1.23-1.8 m/s^2 (12.5-18.4% of Earth). The Galiean moons are large enough to have stable orbit regions around each, but have negligible atmospheres. This makes transport to and from their surfaces easier

Io's surface is composed of silicates, sulfur, and sulfur dioxide. Tidal heating makes it the most volcanically active body in the Solar System. Europa is covered in water ice with a probable water ocean underneath. Ganymede's surface is about 2/3 areas with high water ice content, and 1/3 darker areas with clays and organic material. Callisto has 25-50% water ice on the surface, with hydrated silicates, carbon and sulfur dioxides, and possibly ammonia and organic compounds detected. Surface temperatures of the large moons range from 70-165K, except for volcanic hot spots on Io, and vary mostly by latitude and how close they are to Jupiter, and how muchreflected light they get on the near side.

Outer System

The outer system includes orbits from 2-20 million km in semi-major axes, and 61 known irregular moons. They have much higher inclinations and eccentricities than the inner satellites. Only eight are more than 10 km in diameter, with theremainder being smaller The last two were discovered in 2016 and 2017, so it is likely there are more which are less than 3 km in size to be discovered. 51 of the irregular moons have orbits 20-28.57 million km in size. This is beyond the orbital region we defined for the Jupiter System, but since they are bound to the planet we include them as part of it. Escape velocity from the edge of the region is about 3.5 km/s. Total mass is over 8,000 trillion tons, mostly from **Himalia**, the largest outer moon. Total available solar energy in the outer system is 63 billion TW, or 3 billion times civilization's current energy use.

By size, the largest feature of the Jupiter System is the **Magentosphere**, a cavity in the solar wind created by the planet's strong magnetic field. It extends up to 7 million km towards the Sun, and as far as Saturn's orbit in the other direction. It is filled with highly conductive plasma and contains complex current flows driven by the rotation of the planet's magnetic field with the planet. The magnetosphere traps high energy particles in a belt concentrated between 280,000 and 775,000 km from Jupiter's center. Without shielding, radiation levels are high enough to damage electronics and are lethal to people. This is in addition to the normal solar and galactic radiation present in the Solar System.

Development Projects

We expect development of the Jupiter system to start from the outer edge, and work inwards. Moving closer requires increasing transport energy, and later dealing with very high radiation levels. So it makes sense to start with the outer orbits and moons that are easier to reach. Resupply and support stations can be set up there to prepare for later development of the inner system. The outer areas of Jupiter are also easier to reach from the Main Belt & Trojan areas, which would have previously started development. Equipment and starting supplies can be delivered from these already developed regions.

Together the Galilean moons are over 60% of Mars' mass, and 130 times that of the Main Belt and Trojan region. They differ considerably from each other and from sources in previous phases. The vast and varied source of materials makes eventual development of the inner system worthwhile. It is complicated by high radiation levels and added velocity required as you get closer to Jupiter. The added work required will delay their development until sufficient needs exist. Jupiter itself is much harder to reach than the bodies orbiting it, so any direct uses besides gravity assists is deferred until much later.

Outer System

All of the outer system moons have irregular orbits, are somewhat smaller than the Jupiter Trojans, and their spectra show similarity to some asteroid types. So it is likely they started as captured asteroids. Later collisions fragmented those asteroids to create the current groups with related orbits. The irregular moons have the same average distance from the Sun as the Jupiter Trojans. So reaching them and starting to use their resources should be an extension of previous work in Phase 4D, using the same designs. Although solar energy is weak in this region (3.15-4.3% of the 1 AU intensity), lightweight reflectors should be able to concentrate it to usable levels. Going from the surface of **Himalia**, the largest outer moon, to orbit around it should only take 50-60 m/s, and the remaining moons will need less velocity. So mining and delivery to processing plants should be easy. The Main Belt and Trojan region has much more total material and energy resources than the outer Jupiter System. So the likely reason to develop this area is a step towards the much larger resources of the inner moons.

Inner System

The inner Jovian system has very high natural radiation levels from trapped particle belts. Early development would therefore depend on bulk and active shielding, and remote control from more distant orbits that have lower radiation levels. The unprotected radiation levels affect electronics as well as living things, and would be rapidly lethal to people. We don't know if there is a practical way to permanently change the radiation levels in the long term, due to the strength and size of Jupiter's magnetic field.

Production - Growth of local production would likely follow the usual path of mining first, then seed factories to bootstrap other industries. The smaller outer moons can be an early source of propellants and supplies to support growth, with local sources developed over time. Orbit velocity for Ganymede, the largest moon, is 1,938 m/s, and it is less for the other major moons. An electric catapult can therefore throw cargo directly into orbit for processing in full sunlight, or transport to other destinations. Water is widely available for propellant throughout the system. Rocky and metallic materials may need to be imported from the surrounding regions, depending on the composition of the moons. Large reflectors would be a desirable early product to generate power and heat.

Habitation - Because of the high radiation levels close to Jupiter, we expect most of the habitation in the region to be located farther out. When people are needed, they can occupy heavily shielded habitats and do as much as possible by remote control.

Transport - Transport from previously developed regions would include a mix of electric propulsion, spaceport acceleration, and gravity assists. People and the items they use would travel in shielded habitat modules. High thrust propulsion would be needed for early landings on the large moons, while later transport can use spaceport structures.

Services - Early services include science, exploration, and communications. Later service industries are [TBD].

Jupiter

We don't anticipate much development of Jupiter itself in the near or mid-term. The very high velocity difference from orbit to the atmosphere requires over five times as much energy as escaping the Solar System from Earth's orbit. There are also much easier sources of the gases in Jupiter's atmosphere in the outer Gas Giants and cold regions of Phase 4F. We will keep this heading as a place-holder for the long-term future, and as a spur to finding new concepts.

Phase 5E - Outer Gas Giant Development

This phase includes development of the outer three giant planets: **Saturn**, **Uranus**, and **Neptune** (SU&N). It also includes 103 known moons, a number of which are large, and three ring systems, one of which is famously prominent. There are likely to be more moons that are currently too small and dim to be found. Finally, the phase includes orbits with semi-major axes up to 20, 12, and 12 million km around the three planets. Some of the moons orbit farther than these distances, but we include them in the phase because they orbit their respective planets. Saturn is easily visible to the naked eye, and has been known since ancient times. Uranus is marginally visible without equipment, and Neptune is about six times too dim, so they were only identified as planets in 1781 and 1846. Saturn has been visited four times by spacecraft, including deploying a lander to its largest moon, Titan. Uranus and Neptune have only been flown past once each, by the **Voyager 2** spacecraft. Due to their distance and only a single brief visit each, our knowledge of these planetary systems is less complete than for Saturn

Phase 5E starts after 5D Jupiter System for several reasons. The farther planets need more transport energy to reach, but at the same time solar energy is weaker for propulsion and other needs. Travel times are significantly longer, a number of years by most methods, and temperatures are very cold. It also starts after 4E Outer Interplanetary because these planets are located within that larger region, which must be crossed to reach them.

Region Features

Saturn

Saturn's orbit ranges from 9.04 to 10.12 AU from the Sun, with a period of 29.46 years. It is 95.16 times Earth's mass, with an equatorial radius of 60,268 km, and a polar one of 58,232 km, making it the most flattened major planet. Like Jupiter, it is a gas giant, and does not have a well defined surface. Instead the size is measured where atmospheric pressure equals Earth at sea-level. The upper atmosphere is about 96.3% hydrogen, 3.25% helium, 0.45% methane, and trace amounts of ammonia and other gases. Temperatures range from 84 K (-189 C) at the 10 kPa level to 134 K at the 100 kPa level, and 270-330 K (-3 to 57 C) at the 1-2 MPa levels. There is a complicated set of cloud layers with different compositions at these heights. Gravity varies significantly from 12.06 m/s^2 at the poles to 9.04 m/s^2 at the equator, due to the flattened shape and rapid rotation of the planet every 10.55 hours. These values are approximately equal to Earth's (9.81), but there is no easy way to support yourself in the atmosphere to experience it. The planet is tilted 26.7 degrees to its orbit. Solar flux in the Saturn region is 0.95-1.26% of the 1 AU intensity, except when blocked by the planet, moons, rings, or Titan's atmosphere. Total available solar energy averages 18.6 billion TW in the region.

Saturn has a complex system of 62 known **Moons** and multiple **Rings**. The total system mass is about 1.4 x 10^{23} kg, or 1.9 times Earth's Moon. Seven of the moons are large enough to have become round by self-gravity. The largest, **Titan**, holds over 96% of the mass orbiting Saturn, with an atmosphere containing 95% nitrogen and 5% methane at a pressure of 147 kPa, about 45% higher than Earth. The moons can be grouped into the inner **Regular Moons**, which have low inclinations and are closer than 1.5 million km in semi-major axis, and the outer **Irregular Moons** that are more than 3 million km from the planet and have higher inclinations and more eccentric orbits. The larger moons are generally covered in water-ice, and Titan also has hydrocarbon lakes with methane and ethane. The rings are 99.9% water ice.

Titan has an orbit period of 15.945 days and is tidally locked to Saturn, so the day is the same length. It requires 2.3 km/s to match orbit from outside the Saturn system, and escape velocity is 2.64 km/s. Surface gravity is 1.35 m/s^2. Total surface area is 83 million km^2, or 56% of Earth's land. Surface temperature is 94K (-179 C).

Uranus

Uranus travels between 18.33 and 20.11 AU from the Sun, with a period of 84.02 years. It has 15.91 times the Earth's mass, with a radius of 25,559 km at the equator and 25,362 km at the poles. Uranus is an **Ice Giant**, which is mainly composed of elements heavier than helium. It still has a thick atmosphere of lighter elements, the upper layers being 83% hydrogen, 15% helium, 2.3% methane, and traces of other compounds. Temperature at the 100 kPa level is 76 K (-197 C), rising to about 330 K (57 C) at the 10 MPa level, which is 300 km lower. There are multiple cloud layers in between. Nominal surface gravity is 8.69 m/s^2, with about 0.25 m/s^2 reduction at the equator from its -17.24 hour rotation period. The period is negative because Uranus is tilted 98 degrees to its orbit. Solar flux in the Uranus region is 0.245-0.30% of the 1 AU value when not blocked. Total available solar energy averages 1.67 billion TW in the region.

Uranus has 27 known **Moons** plus a set of narrow **Rings**. The five major moons orbit from 129,000 to 583,000 km from the planet. They are 470 to 1575 km in diameter, with a combined mass of 9.1 x 10^{21} kg, or 12.4% of Earth's Moon. Water ice has been detected on all five moons, and carbon dioxide or carbonate minerals on some of them.

Neptune

Neptune's orbit is quite circular, staying between 29.81 and 30.33 AU from the Sun. It takes 164.8 years to complete an orbit, so it only completed one orbit since discovery in 2011. It is slightly more massive than Uranus, at 17.15 times Earth, and nearly the same size, with radii of 24,764 km at the equator and 24,341 at the poles. Like Uranus, it is an ice giant, with the bulk of its mass being elements heavier than helium, but with a thick atmosphere containing nearly 80% hydrogen, 19% helium, 1.5% methane, and trace gases. The atmosphere at the 100 kPa level is 72K (-201 C), reaching 273 K (0 C) at the 5 MPa depth. Nominal surface gravity is 11.15 m/s^2, decreasing about 0.29 m/s^2 at the equator from the 16.1 hour rotation period. The planet is tilted 28.3 degees to its orbit. Solar flux in the Neptune region is nearly constant at 0.108-0.113% of the 1 AU value, except when blocked by something. Total available solar energy averages 679 million TW in the region.

Neptune has 14 known **Moons** and five **Rings**. Only one moon, **Triton** is large enough to be spherical, at 2702 km diameter and 2.14 x 10^{22} kg mass (29% of Earth's Moon). The remaining moons are only 0.4% of Triton's mass combined. Triton is thought to be a captured Kuiper Belt object because of its retrograde orbit. It is larger and more massive than Pluto, a Kuiper Belt object whose orbit crosses that of Neptune. Triton has a thin nitrogen atmosphere, and the surface appears to be ~55% solid nitrogen, ~25% water ice, and ~15% carbon dioxide ice. It currently orbits 355,000 km from Neptune with a period of -5.877 days. Since it is tidally locked, this is also the day length. Reaching Triton's orbit from outside the region requires 1.8 km/s velocity change, and escape velocity from the surface is 1.455 km/s. Surface temperature is only 38 K (-235 C),

Development Projects

None of the giant plants have accessible solid surfaces, and their atmospheres are mostly hydrogen, making floating systems difficult. They also require high velocities to reach or take off from. So any development of the planets themselves would be very limited. The majority of development would then use the many moons and rings for materials, and either orbital or surface locations in the three systems. Development would likely start as an extension of Phases 4E and 5D, as Jupiter and the inner Centaurs become developed enough to be staging points for sending equipment and supplies to the farther planets

It would then work inwards around each planet in succession. As with other phases, early development would focus on extracting materials to be used in more developed regions. This phase involves large distances and relatively difficult environments, so a significant amount of research and development will be needed before it can start. Prior development in multiple preceding phases should allow enough time and experience for that to happen.

Production

Due to low solar flux, either large, lightweight reflectors would be needed to bring it to useful levels, or alternate energy sources like nuclear. Orbit velocity around Titan is 1.86 km/s, making gas mining particularly easy. The difference between orbit and equatorial rotation velocities for SU&N are 15.2, 12.5, and 13.9 km/s. This is possibly low enough for mining their atmospheres from orbit.

Helium-3 has been proposed as a low radiation fusion fuel. Fusion in general has not yet been solved, and the He-3 - He-3 reaction is harder than the Deuterium-Tritium one that is the main target of research. If the technical issues are overcome and a need develops, Uranus and Neptune have the highest ore concentrations of Helium in their atmospheres, and thus the He-3 isotope. At the same time, fusion reactors would be available to make trips to these planets in reasonable time. Such uses are far enough in the future that technology is likely to change dramatically in unexpected directions. So it is too early to make plans for projects like this, but we can note the possibility for long-term development.

Habitation

All three of the outer giants have magnetic fields which create trapped high energy particle belts. They are weaker than those around Jupiter, but likely still a hazard to unprotected people and electronics. Since Jupiter is a worse case, designs should already be available from the previous phase, but more work is needed to understand the severity in different parts of the respective outer regions. Even outside the radiation belts, some solar and cosmic radiation is present generally throughout the Solar System, requiring a moderate level of protection.

Transport

Direct transfer ΔV departing from Earth's orbit to the SU&N regions are 10.3, 11.3, and 11.7 km/s, the latter not far from Solar System escape at 12.3 km/s. To match interplanetary orbit with the regions when arriving requires 5.45, 4.65, and 4.05 km/s. Orbit velocities at the edges of the regions are 1.377, 0.695, and 0.755 km/s, and low orbits around the planets are at 25.0, 15.0, and 16.5 km/s. So destinations within the region require a variable additional ΔV. Reaching the regions from Earth can use planetary gravity assists, and all three of outer giants and their larger moons have enough mass to help at arrival and later orbit changes. The larger moons can probably support stable enough orbits for orbiting spaceports, which can lower the velocity to get to and from the surface.

Low orbit periods for SU&N are 250, 180, and 155 minutes, and are 2.89, 3.44, and 3.16 years at the edge of their regions. Transit time between points in the region will vary beyond these limits depending on the transport methods used, and needs for orbit phasing and inclination changes. Direct transfer orbits from Earth will take 6, 16, and 30 years to reach the regions. These value are high enough that for people and equipment you would want to do something useful during the trip, or use faster transport systems. For bulk cargo you might use slower but more efficient routes with gravity assists.

The power source for outer system transport is still to be determined. Nuclear fuel has an energy content of ~80 TeraJoules/kg, while solar panels near Earth produce around 80 GigaJoules/kg. The relative output then depends on the conversion efficiency of nuclear reactor fuel energy to useful output, and the mass ratio of the reactor system to the fuel load. Fissionable elements are relatively rare compared to silicon used for solar panels or aluminum/magnesium used for concentrating reflectors. So for large-scale energy demands, the solar sources have better material supply. However, in distant regions where sunlight is weak, nuclear approaches may have the advantage. Although reactors emit harmful radiation, most parts of the Solar System are filled with harmful radiation anyway. The same shielding that protects people and equipment can do it from both sources.

Services

Early services are likely to include science, exploration, and communications. Later service industries are [TBD]. Round-trip communications times from Earth to the SU&N regions are 2.19-3.12, 4.78-5.87, and 7.96-8.71 hours. If communications are to be maintained when Earth is in opposition, a relay link will be needed, which slightly increases the maximum time.

Section 4.12: Phase 6 - Interstellar Development

Phase 6 is the last major phase of our program. The key difference that warrants a new major phase is the extreme distances involved. This mostly breaks the ability for timely delivery to and from the region, and round trip communication with the Solar System is measured in years to decades. Expansion of civilization to these regions would require high self-sufficiency in transport and good enough production systems to enable growth without much assistance. It is far enough in the future that we can only speculate about development in very general terms. We currently divide this major phase into three parts: 6A, 6B, and 6C. The first two cover the interstellar regions between stars to a distance of 20 light-years (LY), and stellar systems, including the regions their gravity dominates, to the same distance. The third covers everything beyond 20 LY, but for now that is mainly a place-holder. This section gathers our early ideas about these three phases, pending further concept exploration work.

Interstellar space, the cold regions between stars, is not much different from the environment of the outer parts of the Oort Cloud in Phase 4F. We know very little as yet about equivalent cometary clouds around other stars, or free-floating objects not attached to stars. We have better information about planets and dust disks around other stars. Their parent stars tell us where to look, and the stars themselves provide some data about the planets from Doppler shifts and transits. Disks are visible mainly in the infrared, and are found around younger stars. The number of discovered planets is growing rapidly, from none before 1988 to 3500 confirmed and 4500 candidates by the end of 2017. We expect there are also many smaller objects in systems with planets, but today we can only infer them indirectly.

A future method to inventory these smaller objects is using the Sun as a gravity lens. It brings objects into useful focus around 800 AU from us, in the Scattered Disk region. Placing telescopes directly opposite a star of interest would allow much more detailed observations than otherwise possible, because the 2 million km effective diameter of the lens allows extremely high resolution. Putting instruments at such high distance from Earth and getting data back will not happen until Phase 4F, which is well in the future.

Since even making an inventory of what resources are available requires advances in technology, plans to use those resources for development are even further off. They will depend on experience gained in earlier phases and require great improvements in transportation methods.

Phase 6A - Nearby Interstellar Development

The Nearby Interstellar region is the next in distance after the Oort Cloud portion of Phase 4F. It begins with orbits whose semi-major axis a = 100,000 AU or more from the Sun, where our star's gravity is no longer dominant. For design purposes we set an arbitrary outer boundary of 20 LY from the Sun. If we can gather supplies and rebuild our equipment at that distance, then later projects can travel farther in increments of 20 LY using the same designs. The **40 LY diameter sphere surrounding our Sun** makes up the nearby portion of the **Solar Neighborhood** **(Bovy, 2017)**, which extends to 250 LY. This in turn is a small is part of the 100,000+ LY diameter **Milky Way** galaxy we belong to. We are about half-way from the core to the rim of the galaxy, near the central plane.

This phase covers the spaces between stars, so it excludes star systems, the objects which orbit them, and the orbital regions of gravitational dominance surrounding them. The Phase 6A region's volume therefore resembles the solid portion of Swiss or Emmentaler cheese (Figure 4.n-5), with holes around each star system not included. Natural or artificial objects which are moving fast enough they are not tied to any star, and are more than half the region boundary from any star (i.e. >50,000 AU in the case of the Sun) are counted as interstellar. Otherwise they are considered temporary members of a stellar system.

Region Features

The open space between stars includes a number of components. The most massive of these include **Sub-Brown Dwarfs**, which formed the same way as stars by the collapse of a gas cloud. They do not have enough mass for deuterium fusion, and are therefore not stars. Their masses range from 1-13 times Jupiter's (M_J). The lower bound is set by not having enough mass to collapse, and the upper bound by sufficient mass to initiate fusion, making them stars. Several such free-floating objects have been detected. They do not orbit a larger brown dwarf or regular star. The other route to forming large interstellar objects is from a planetary system which forms around around a star. Later gravitational interactions can eject some of the objects into interstellar space. The largest lost objects are called **Rogue Planets**. Their mass can range from the same upper limit as sub-brown dwarfs (13 M_J) down to a lower limit where they don't have enough mass to assume a round shape. This is somewhat below 1000 km in diameter, depending on composition, at which point they are no longer considered planetary size. Rogue planets are distinguished by having a higher concentration of heavier elements than sub-brown dwarfs. This is due to more of the heavier elements tending to condense into planets, and the lighter ones tending to be blown away by the parent star.

Simulations of the history of the Solar Nebula (**Shannon, 2014**) indicate that about 80% of the original small bodies within 40 AU of the Sun were ejected into interstellar space. With over 3,500 confirmed **Exoplanets** by late 2017, we now know that formation of planetary systems is common around stars (see

NASA Exoplanet Archive). So if the same ejection process happened for other stellar systems, then interstellar space should be filled with a large population of objects from many stars. The largest of these objects would qualify as the rogue planets previously mentioned, but their size distribution should continue down to dust-sized particles. We define dust particles as those less than 1 mm in size, and **Meteoroids** as those from 1 mm to 1 meter. Between 1 meter and planetary size, we class them as comets if they are icy, and asteroids if they are rocky. The different compositions come from where they originally formed around a star, and later events in their history

Figure 4.n-5 - Emmentaler cheese.The solid portion represents the volume of the Nearby Interstellar region.The holes represent the regions around stars, which are assigned to Phase 6B.

Only one sub-planetary **Interstellar Object** has been discovered so far, in late 2017. **1I/'Oumuamua** is only about 160 meters in size, and apparently of rocky composition. By chance it happened to be 0.2 AU from Earth when discovered, making it close and bright enough to be detected. It is moving 26 km/s above solar escape velocity, so was never bound to the Sun. We count it as a temporary member of our Solar System for the next 9,000 years, until it reaches 50,000 AU distance.Due to the relative velocities of their parent stars, and the age of the Milky Way, objects like this one, that are currently within 20 LY, could have started anywhere in the galaxy, and even from outside it.

A few of the most massive non-stellar objects, in the Sub-Brown Dwarf range, have been detected by their infrared glow. Smaller objects will have rapidly cooled to ambient interstellar temperatures, and would have nearly no light reflected from nearby stars, making them extremely difficult to detect with current instruments. Therefore the population of these smaller objects is nearly unknown at present, and only roughly estimated from losses by our own Solar System. A possible future method for finding them is to use natural gravity-focused light from stars, or artificial lasers, to scan around a region looking for reflections. Moving the scanner along interstellar paths would then build up a map of object locations. More investigation of this concept is needed to determine if it is feasible, and other detection methods should be pursued.

In

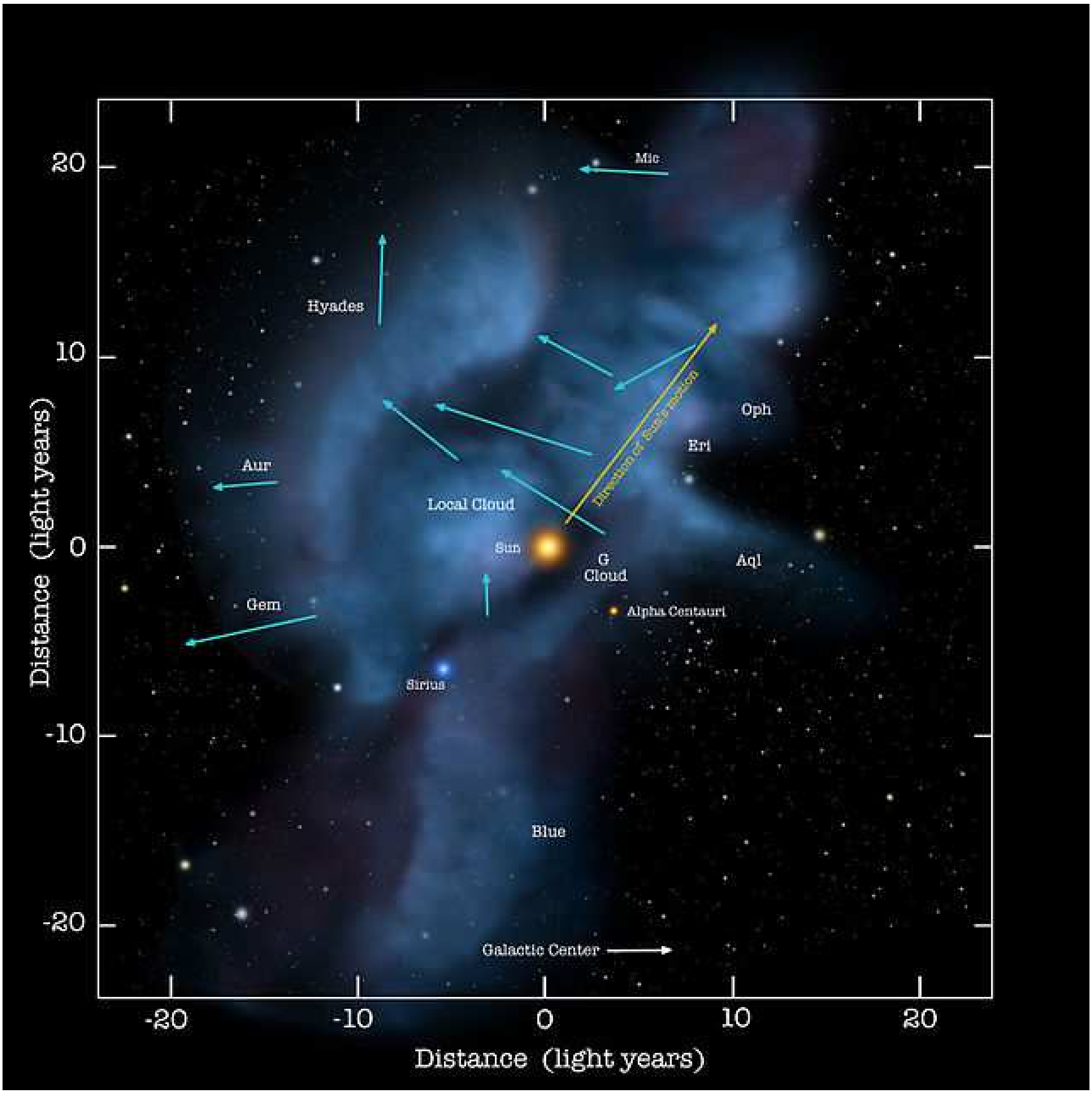


Figure 4.n-6 - The Motion of the Sun and nearby interstellar clouds.

addition to the objects larger than 1 mm, the **Interstellar Medium** between stars contains gas, dust, charged particles, magnetic fields, and electromagnetic radiation. The density and temperature of the medium varies by location. There is also the hypothetical **Dark Matter** and **Dark Energy**. We don't yet understand what the "dark" components are or how to use them. They are of scientific interest, but they can be ignored as far as our program is concerned. The Sun is presently moving through a region of slightly higher gas density called the **Local Interstellar Cloud** (Figure 4.n-6). It will continue to do so for the next 10-20,000 years. The local cloud has a gas density of about 0.3 atoms per cubic centimeter, or 1 gram per 564 km cube. This does not include other components of the interstellar medium, or any larger objects that may be present.

Stellar energy sources are too small in the region for practical use, except possibly along lines of gravitationally focused starlight. Ambient temperatures will mostly be close to the cosmic background temperature of 2.7K. Travel time is many years with known technology, and depends on future improvements to reach useful engineering time scales. Round-trip communication time will range from 3 to 40 years from Earth, and up to 80 years across the region. Stellar radiation is generally not a factor in this region, but cosmic radiation still is.

Development Projects

We don't yet know enough about the material and energy resources in the region to do more than speculate about development projects. We think it is possible the region will be used for fast travel by self-contained vehicles on their way to other star systems, or for slow travel by permanent colonies, who use local resources as they go. Stationary locations don't mean much in a region where everything from stars to gas clouds are in motion relative to each other. Even if you stay put relative to the Sun, other things will still move past you. The great distances from the Sun and other stars are likely detach interstellar industry from regular trade with the rest of civilization. Science, exploration, and seeding interstellar colonies are possible future activities.

Production

Concepts for mining materials from the region are deferred for now, until better data is available on what is present. The production functions we consider for now are aboard transport vehicles and colonies. Trips with even advanced transportation will take years. So fast vehicles will still likely need to do maintenance and repair, either from spare parts and supplies, or to produce new items from wastes and scrap. These technologies should have been developed in the previous phases around the Sun. Permanent interstellar colonies don't reside around one star. They travel between stars, and either stop to gather supplies, or collect them while in motion. Such colonies may be planted on or around existing large bodies, and mine them for resources. Changing the course of a large body would be difficult, so the colony would have to accept the existing trajectory, migrate to a different body, or spawn new colonies which then follow their own courses.

By their nature, isolated interstellar colonies would have to produce most of what they need locally, with perhaps some deliveries by fast transport. The first such colonies would be built somewhere in the Solar System, then placed on their chosen course once functioning. They may evolve from Oort Cloud colonies, whose environment is not significantly different from interstellar space. Any type of production or other function will require energy. Two possibilities are gravity-focused beamed power from local stars, and nuclear energy, either fission or fusion.

Habitation

Habitation designs should not be significantly different from those developed for the outer regions of the Solar System, since the environments are similar.

Transport

Interstellar transport can be divided into slow and fast types. The slow type is on the order of stellar velocities (5-500 km/s) using large habitats with large material reserves and nuclear or beamed energy sources. These habitats can potentially mine cometary clouds around stars and rogue objects between stars. When they get close enough to a selected star, they can enter orbit and travel with it, either permanently or later set course for a new destination. At any time, such habitats may build and spawn additional ones. Travel times between stars at slow interstellar speeds would be 3000 years or longer.

Fast interstellar is dominated by the higher energy required to reach velocities greater than 500 km/s, and shorten the time to a destination. Possible transport methods are discussed in Part 2 of this book. High energy candidates include fusion powered engines and beamed power from local stars. Rather than a large habitat with a full range of civilized activity, fast vehicles operate more like ships on Earth, with a crew dedicated to reaching a destination and maintaining operations.

Services

Possible future services located in interstellar regions include science, exploration, and communications relay.

Concept Details

Slow Interstellar

Based on past exposure to fictional works, most people assume that a "starship" will be a futuristic shape with big engines on the back. Instead, imagine colonizing a long period comet, one of the ones that came from the Oort cloud, and is heading back out there. Comets are made of a mix of ices (water, methane, ammonia, CO_2, etc) and rocky materials. If there are not enough metals, then a metallic asteroid can be matched to its orbit. You then build your colony mostly out of the materials already there. Comets range in size up to 50 km in diameter, and there are a large number of outer system bodies more than 100 km in diameter. A 50 km comet contains about a 1500 year supply of Earth's total mining output, which should be more than enough to sustain a colony with recycling.

The Oort cloud is many times the distance of the Earth from the Sun, and the velocity needed to get the comet to leave the Sun and head for another star is very small. All the ices have some amount of hydrogen, and thus deuterium, which means if you know how to build fusion reactors, you have power for a very long time. Rather than a sleek ship, your "vehicle" is a city attached to or built around a comet core, which over time is converted into necessary items for the colonists. It will be a long trip, but you have a large amount of space to live in, and the colonists can make occasional side trips to other comets as they go past to get additional supplies.

There are an estimated trillion comets in the Oort Cloud around the Sun, and some will be along your existing route, more or less. The average spacing is something like 6 AU, about the distance to Jupiter. At the slowest interstellar velocity, 5 km/s, you would pass one per 6 years on average. So there are plenty of mining opportunities, and you can in theory seed other comets as you pass by with new colonies. If some people feel like it, they could head back to the Sun, the velocities are low enough to do that. When you reach the edge of the Oort Cloud region, you can continue this sedate interstellar travel by using objects between the stars, and the cometary clouds around other stars. The requirements for this kind of slow star travel include fusion power, and knowing how to build permanent habitats in space. Both of these should have been developed in previous phases. Another requirement is a way to detect small objects along the travel path, in the absence of starlight to illuminate them.

Starlines

Gravitational lensing occurs around every massive object. In fact, measuring the bending of light during a solar eclipse 100 years ago was the first proof of relativity theory. For the Sun, the light bent from all sides comes to a focus at distances greater than ~540 AU. The focus is not to a point, but rather a radial line. This is because photons that miss the edge of the Sun by a larger distance are bent less, and thus focus farther away. Every star in the sky therefore produces a line of focused light on the other side of the Sun, and so we call them **Starlines**. Every other star in the interstellar region also produces a pattern of starlines surrounding it. This forms a network of lines of light filling interstellar space. If the starlight is sufficiently well focused, it may prove useful for interstellar power and propulsion.

Phase 6B - Nearby Exostellar Development

Interstellar travel and planets around other stars have been explored in **Science Fiction** for a number of decades. The authors of such works can assume whatever transportation methods and planetary environments are needed for their stories. Engineers considering development projects can draw their ideas from fiction, but are limited to actual technologies and real places to implement their plans. This phase of our program covers development of the regions around stars other than the Sun. It logically follows Phase 6A, which is concerned with the regions between stars, since we must travel through such regions to reach other stars. It also follows all the earlier phases which are dedicated to developing various parts of our Solar System. Like Phase 6A, we limit this one arbitrarily to within 20 light-years of the Sun. This is a large enough sample of stars and their attendant systems to identify design requirements for them, and is a large enough range of distances to identify whatever problems that will cause. Development at distances beyond 20 LY are reserved to the last current program phase, 6C - Farther Interstellar.

For our purpose, we identify as stars any object large enough to undergo nuclear fusion, past, present, or future. This includes brown dwarfs who only undergo deuterium fusion, larger stars that mainly undergo hydrogen fusion, and stellar-mass remnants like the six **White Dwarfs** which are closer than 20 LY. A stellar system may include one or more stars bound by gravity, and all the attendant objects and material which orbits them. We define an orbital region surrounding each stellar system extending from the center of mass, to orbits with a semi-major axis of 100,000 AU times the square root of the system mass in units of the Sun's mass. This is a region where the local system gravity dominates the rest of the surrounding galaxy, and they are able to retain cometary clouds. If two systems are near each other the orbital regions may overlap. In that case we draw a boundary between them where their gravity is equal. The stellar systems and their orbital regions are bubbles set within the larger Nearby Interstellar region filling the space between them. We separate Phases 6A and 6B because stellar systems have more in common with the earlier phases around our Sun than the spaces between stars.

Region Features

Possibly the most significant feature of the region is the stellar systems which make it up are all in relative motion to each other, with an average velocity of 50 km/s relative to the Sun. This motion is in addition to the general rotation of the Milky Way galaxy, which is about 225 km/s in this area. There are about 105 **Nearby Systems** within 20 light years, including our Sun. This list should be nearly complete. Given their average velocity, they will take 120,000 years to travel 20 light-years, so the nearby population will change about every 1150 years on average. Current transport methods require much more time than this to reach 20 light-years. So future plans for the region should take into account the motions of the stars and changes in the nearby population.

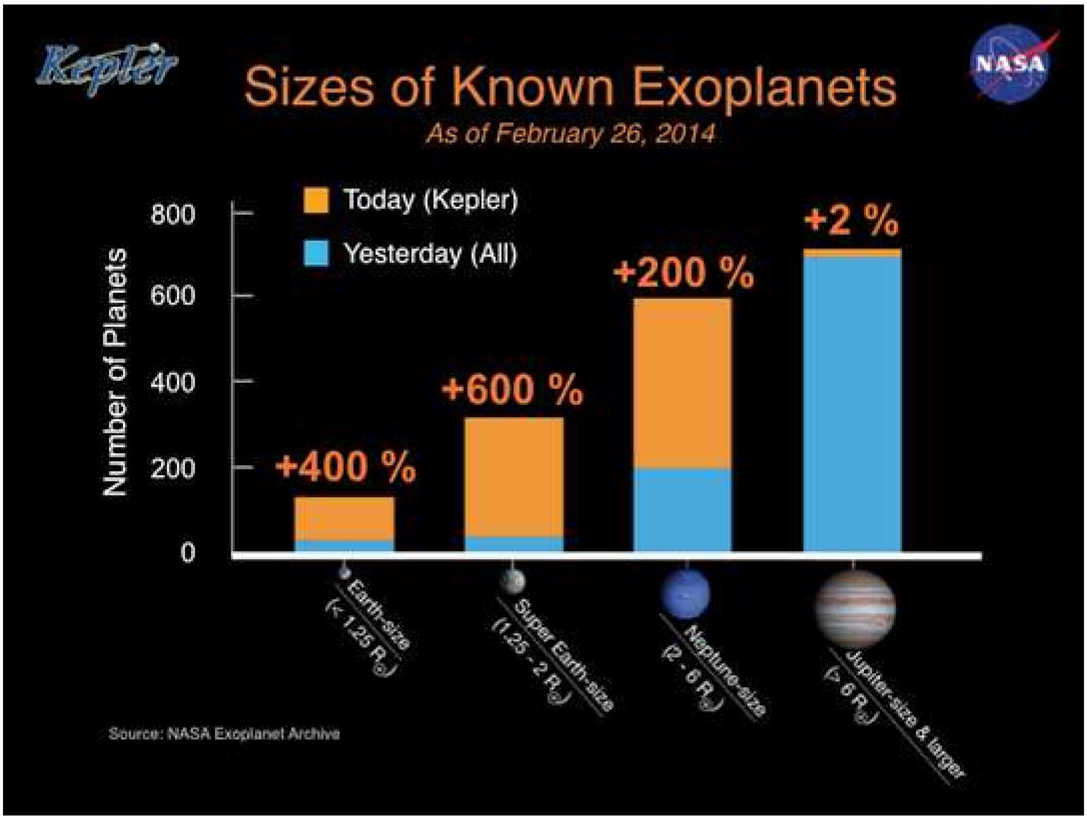


Figure 4.n-7 - Number of known exoplanets by size as of 2014.

As of 2017 there are about 31 confirmed exoplanets outside of our Solar System, and within the 20 LY limit. Our detection methods are still biased towards larger planets (Figure 4.n-7), with nearly all confirmed planets larger than the Earth's mass. The size distribution in our own Solar System indicates the number of objects increases as their size decreases. This makes it likely that more smaller planet, and many more objects smaller than planets exist around other stars and await discovery. We expect the count of nearby planets to increase over the next few years as our instruments improve. There are also two known nearby **Circumstellar Disks** around **Epsilon Eridani** and **Tau Ceti**, which are detectable by their dust component. It is likely comets and asteroids and some larger objects also exist around these stars but are not yet detected. Those would be in addition to the Earth-mass or heavier planets already discovered or suspected.

Development Projects

We have a reasonably complete list of stars to 20 LY, because they are bright and nearby. We don't yet have a complete list of planets and know nothing of smaller bodies aside from dust particles when they form disks. Therefore concepts for developing the Nearby Exostellar region must remain speculative and preliminary at this time.

Production

As mentioned in the introduction to Phase 6, we would like to observe nearby stars and their surroundings by using the Sun as a giant gravitational lens. This should enable us to inventory their resources in much more detail than near-term telescopes. Following that would be sending smaller robotic probes to more closely examine whatever is found around these stars. Once sufficient information is available we can plan how to start production in the region. Self-expanding production has been a theme throughout our program, and should be well developed by this phase. So we would expect to also use that approach around other stars. An open question is whether to start production with a robotic seed factory sent ahead of people, or begin with a larger system that arrives with the first people and is already supporting them. We probably can't answer that question until more experience is gained around our home star.

Habitation

Habitats for people and other living things should be mostly similar to those developed for regions around the Sun. Differences in temperature, radiation levels, and stellar spectra may require some modifications to previous designs.

Transport

Transport between stars is covered above under Phase 6A. Travel within a given stellar region should be able to use similar technologies as developed around the Sun. One difference is the brightness and temperatures of stars vary, so stellar energy based transport would need modification to account for this.

Services

Due to extreme distance, the service activities we see for now are confined to science, exploration, and seeding independent colonies.

Phase 6C - Farther Interstellar Development

Our last defined phase is a place-holder to cover the remainder of the accessible Universe. It begins 20 LY from the Sun, and extends as far as transport methods make it possible to reach. Current and near-term transport is very far from able to reach such distances. So the only work we assign to this phase for now is transport improvements. Other work is reserved to some point in the future. This includes defining additional phases if the need arises.

Part 5: Design Studies

Introduction

In the previous part of the book we give an extended example of a program for Human Expansion. The narrative there is intended to show why complex programs exist, how they are organized and designed, and how they evolve. For that purpose, a summary of design studies and decisions based on them is sufficient. In this fifth part of the book we include the design studies in their full detail. This is to show by example how the analyses, calculations, and decisions are done. We also include incomplete work-in-progress studies. Readers are encouraged individually or in teams to add to and improve these studies. This gains skill and experience applying the methods described earlier in the book, and in working in teams, which is how most real projects are carried out. Contributions to studies can also serve as resume items when looking for paid work. which are too detailed or too incomplete to include in the main narrative of the book. Any added results from these studies will be incorporated into earlier parts of the book. As an electronic book, it can be constantly improved, rather than the somewhat static condition of paper textbooks.

The list of studies below is organized first by level of study completion, then by expected order of use within the program.

Full Studies

We use the term "full" rather than "completed" because no engineering work is every truly complete. As time passes, assumptions made in the study will be affected by real-world changes, and new technology and ideas will get developed. As a consequence, the study could be updated or revised to reflect these changes. For an actual engineering project, however, you must reach a conclusion or make a decision, and proceed to more detailed design or production. This section includes studies which are sufficiently complete that such conclusions or decisions can be made from them. At present (Oct 2012) none of the studies have reached that point.

Studies in Work

These are studies which have been started, but have not reached completed or full status. They are given their own sections, with as many pages as needed by their length, and approximate completion status is noted here.

- **Conceptual Design for Human Expansion (Status: approx 20%)** - This study carries through a first stage analysis and design of the program as a whole. The principal purposes are (1) to determine if such a program is desirable over existing programs, and (2) establish one or more baseline designs for the next stage of work if the program is found desirable. Because this study looks at the whole program, it is the first one to be worked on.

- **Seed Factory Concept Development (Status: approx 10%)** - This study attempts to formulate concepts for a "Seed Factory", a starter set of equipment that both outputs more equipment to expand itself and also useful products. It is part of a separate Wikibook on that topic.

Study Ideas

These are ideas for future studies which have not been started yet. They do not have their own section yet, but are merely described here.

Starting Location Trade

The two nearest objects in velocity terms beyond Earth orbit are the Moon and Near Earth Objects (NEOs). If you want to obtain materials or set up production, the question is which to start with. The options are Moon first, NEO first, or both in parallel.

Composition

The composition of the Moon and NEOs is different. The Moon does have some Carbon in it, around 100 parts per million. Chondrite asteroids are approx 0.4% carbon (4000 ppm). So given 40 time richer "ore", and the ability to run your processing plant 100% of the time vs 50% on the Moon (lunar night), you likely want to mine an asteroid to get the carbon to *build* a space elevator. Like for terrestrial mining, location of the mine and richness of the ore determine where you want to mine.

Access

Discarded Studies

Some studies will turn out to be unnecessary, merged with other studies, or otherwise determined to not be useful. They are listed here to have a record of work done previously. This is different than a completed design study for an option not chosen. In that case the study was useful in making the choice among alternatives.

No discarded studies are in this category yet.

Section 5.1: Conceptual Design for Human Expansion

Conceptual Design Stage

The first stage in the life cycle of a well-designed program is Conceptual Design. The goals of this stage are to identify the needs to be satisfied, the selection criteria to be used, an initial concept for the major system elements, and how they will be built, operated, maintained, and disposed. The alternatives to building a new system are to do nothing or to continue with existing projects and programs. We therefore assess our system concept against those alternatives and decide whether to go further or to stop. We do this using the selection criteria previously chosen.

For our example program we give the motivation in terms of general goals first, with potential reasons to change from current space programs. Not every goal will have the same importance to every person or organization. As long as a sufficient number of goals apply to them, they would have a reason to support the project. Following the goals we describe our design approach in general terms. The rest of Part 4 will go into more detail of the design process and results.

[text inserted from elsewhere to be merged]

Conceptual Design

Conceptual Design is the first stage in the life cycle of a program or system. We determine needs and goals to be satisfied, what criteria to use in selecting among designs, and a design concept for the major parts of the program, including how they will be built, operated, maintained, and disposed at end of life. We compare our new program to existing programs using the same selection criteria. If the new program is "better" by these measures, we recommend going forward to the next design stage. If it is not better, then we stop and wait for circumstances or technology to change, or try to devise alternate concepts. The full details of our reasoning, calculations, and decisions are too long to include in our discussion here. We will summarize them here, and refer readers to the full **Human Expansion Design Study** for the details.

The general process flow in Conceptual Design starts with general description and goals, which are noted in the next section. Following that we perform an analysis step to develop these goals into more detailed and quantitative statements. At present this has been done for the program requirements and evaluation criteria. The functional analysis to define more detailed program elements has been started, and later parts of the Conceptual Design process are incomplete. Candidate designs to fulfill the program elements are collated in the remainder of Part 4. They should be considered as pieces which might fit into the overall program, but have not yet been linked, or completed individually.

Program and Customer

We begin by identifying who is the **Customer** for our program and what their needs are. Any proposed program involving many people needs sufficient motivation for why to do it instead of the alternatives of doing nothing, or continuing with what is being done now. If you are building a system entirely by yourself, then "I want to do it" is enough, but for larger systems you typically have to convince end customers and developers of its value. The motivations can be couched in terms of general goals, or financial benefits, or simple efficiency, or even "I have a gun, do this or else" for totalitarian societies, but they must be there in some form. When the end customer is unaware of a program's value, they need to be informed about it, preferably not at the point of a gun. Thus how to format and deliver the results needs to be considered along with the technical results.

Customer vs Designer - Customers can be internal or external. The former case is when the same organization is both the user and designer of a system. The latter case is the more common one for large systems, where design specialists develop systems for another entity than themselves. For our future program then we must ask who is setting the program goals and is thus the customer, and who is doing the design work? For now both are actually the editors of this page, but we will assume a non-profit foundation will continue the initial design work using open-source methods. As a foundation they would set goals for the benefit of civilization as a whole, and publish the results of their work publicly. Assuming good enough results are reached, then fund raising for later research, start-up of commercial entities, or promotion of government funding would be used for later stages. The ultimate customer is therefore civilization as a whole, and the designers are part of an open-source non-profit foundation. At present this is the Wikimedia Foundation editors of this open source book, but that may transition to a specialized organization later.

Customer Acceptance - Some people or groups will be negatively affected by a new project, in particular those associated with older existing projects. Other people simply do not like change, or are averse to the risks of new or untried methods. Yet others have a preference or aversion to particular designs separate from their technical merits. Finally, some methods are simply new and unfamiliar and not considered for that reason. A system designer has to understand these human elements and respond to them, and not assume that the best technical answer will be accepted just because it is the best technical answer.

Program Constraints - Beyond technical design, there are also outside factors, such as available funding or restrictions on technology transfer, which affect the course of a program. These outside factors need to considered in addition to the technical and human ones.

Program Goals

Human civilization desires, as the sum of individual desires, to survive over the long term, with lower risks individually, and to flourish materially and socially. Therefore we state the following individual program goals to satisfy these desires. We group them into those that benefit civilization in general, and then those that apply to individuals or groups of smaller size. By the end of the conceptual design stage, these will be refined to final goals and specific numerical objectives.

Civilizational Goals

Our earthly civilization should be motivated to visit, develop, and expand into space for a number of reasons. These reasons can stated in terms of individual goals. A future program or programs ideally can satisfy more than one of these goals at once:

- ***Improve Life on Earth*** - Attempting new and challenging things tends to discover new knowledge and develop new technology. Such developments can first be applied on Earth to improve our life here, particularly in the areas of energy resources, automated production, and closed ecologies. With experience gained on Earth we can then use these developments in space.

- ***Understand the Earth Better*** - We learn more about how the Earth works by looking at it's current environment in space, because the Earth is not isolated in the Universe. We also learn from looking at other examples of planets and environments and how they evolve. Except for the one part per billion of human kind currently in orbit, all of us live on the Earth, it's a good idea to understand it better.

- ***Reduce Hazards from Space*** - There are hazards in space such as solar flares and asteroids which can affect us here on the ground. In order to prepare for or prevent these hazards, we first have to know their magnitude and characteristics, followed by developing methods to prevent or deal with the hazards.

- ***Increase Biosphere Security*** - There is only one biosphere right now that we all depend on, and it is prone to natural and man-made variations (ice ages, CO2 changes, volcanoes). Even the International Space

Station relies on food and other supplies which come from Earth. Observing from space helps us understand how the Earth varies better. In the long terms we should also set up backup biospheres for survival reasons in case something catastrophic happens to the only one we have.

- ***Expand Material and Energy Resources*** - Civilizations require materials and energy to function, and there are vast material and energy resources beyond the Earth. We could move heavy industry and population off the Earth to reduce our impact to the planet.

- ***Long Term Survival*** - In the long term the Earth is doomed as the Sun continues to get hotter over its life, and eventually turn into a red giant and swallow it. Long before then, it will become a poor place for life because the Sun increases it's luminosity over time, eventually overheating the planet. Our choices are planetary scale engineering to deal with the eventual heat, or moving elsewhere. Either way involves space in some way.

Localized Goals

In addition to goals which apply to civilization as a whole, there are motivations which apply to particular individuals or groups. These more localized goals can overlap with the more general ones listed above:

- ***Increased Choice and Freedom*** - Freedom of choice is restricted in many ways by an occupied and relatively crowded Earth. You cannot just set out and start your own community with your own rules for living, because all the land area of the Earth is already claimed by some government and functions under their rules. Dangerous or dirty experiments, whatever their scientific value, have to be restricted, and some experiments, like terraforming, simply cannot be done on the Earth. At least to start with, space is unoccupied and will not be crowded for a long time, so earthly restrictions are lifted. Space will impose it's own restrictions due to environment and resources, but they will at least be different restrictions. This widens the total range of available choices.

- ***Increased Opportunity*** - Unlike Earth, where almost everything is already claimed by someone, there are many unclaimed resources in space. Lack of acquisition cost from a previous owner means a wider opportunity for gaining wealth, but of course not a certainty of it. The ratio of available space resources per person will start very high, partly because there will not be very many people at first, but also because the absolute amount of matter and energy available in space exceeds that on Earth by a wide margin. This high relative availability of resources creates new opportunities for those who wish to exercise them.

Program Benefits

In theory the above goals could be met by existing space programs. So in addition to stating our goals and the reasons behind them, we have to demonstrate why a new system is better than just continuing what is being done now. This is the intended result at the end of the Conceptual Design stage. Where we are now, which is the start of that stage, we can only list the ***potential*** benefits identified so far. Additional ones may be identified later, and some of those listed below may not prove favorable once the concepts are developed further.

New Systems Can Lower Cost Dramatically:

If we stay with existing programs, we will get existing planned costs. The planned costs include future projected cost reductions of systems already in development. A new system has the potential to reach even lower cost. The following argument establishes the potential gain is very large. Therefore new systems should be examined to find out if they can actually reach some part of that potential. This can only be done with a sufficiently detailed design that considers all of the cost elements.

- ***Existing program costs are much higher than what is possible:***

With all the many billions of dollars spent by governments and private industry on space projects, we have only managed to get about 1000 useful satellites and a half dozen people working in space. This is because past programs had launch cost measured in their weight in precious metals. For example, the Atlas V is listed as delivering 20 tons to Earth orbit for US $110 million. The cost of

$5,500/kg comes to $171 per troy ounce, vs a mid-2012 price of $1600 for gold. The new **Falcon 9** commercial rocket reduces that somewhat to 13.15 tons for US $54 million, or $128/oz ($4,100/kg), and the upcoming **Falcon Heavy** is quoted by the manufacturer at 53 tons for US $83 million, or $49 per ounce ($1,566/kg). This compares to an average of $30/oz for silver in the first half of 2012

The raw energy cost to get to Earth orbit, which is about 33 MJ or 9.25 kWh/kg, works out to about $1/kg at typical retail cost for electricity. This is slightly higher than the price of **Carbon Steel** as of 2012, or roughly the retail price of potatoes per kg. Efficient shipping to orbit should in theory run several times raw energy cost, or roughly the cost of cheese or priority mail packages. The ratio of upcoming launch costs and theoretical cost is still about 400:1, indicating there is much room for improvement. As long as launch costs are measured in their weight in precious metals, rather than cheese, it should be evident not much will be done in space. One billionth of the world's population being in orbit counts as "not much".

- ***High costs are driven by launch technology:***

Conventional rockets, which carry all their own fuel, cannot get really cheap and efficient because the energy to reach Earth orbit is about twice the energy in chemical rocket fuel. Therefore you have to burn a lot of fuel to get a smaller amount of fuel halfway to orbit, which in turn you burn to put an even smaller amount of cargo into orbit. So after subtracting fuel, fuel tanks, and rocket engines, about 3% of what sits on the launch pad is the delivered cargo. Even worse is that the hardware is expensive aerospace-grade hardware and is often used only once and tossed away. By comparison, airplanes use the same kind of aerospace grade hardware, and so cost about the same $/kg to build. But being more efficient (they get oxygen from the air, rather than from a tank) they carry a higher percentage of cargo, and the airplane is used many thousands of times, thus making it dramatically cheaper per trip.

- ***Launch cost and lack of resource use drives high total cost:***

The high cost of everything else in space is partly because of the high cost of getting to orbit, and partly because everything you need in space is currently brought from the ground. Since shipping is so high, designers spend a lot of effort making the cargo lighter by spending more time refining the details, and by using more exotic, but stronger, materials. That extra effort makes the cargo more expensive also. Since you optimally spend money making it lighter to the point where each kg saved costs the same as launch cost, cargo cost tends to be proportional to launch cost. If you lower the cost of launch, then it will also tend to make whatever you are carrying cheaper (and heavier).

The other important cost factor is the practice of bringing everything from Earth, and thus paying shipping on all of it. If you could get some of needed supplies from space and recycle more, the amount that has to be shipped to do a project or mission could be cut drastically. There are plenty of energy and material resources in space if we learn to use them, and for many of them the shipping energy is dramatically lower than from Earth. The combination of launch cost, cargo cost, and bringing it all from Earth results in very high total project cost.

- ***A potential improvement of 20,000 times is possible:***

Even an ideally efficient launch system will still have costs for development and operation, therefore will not reach raw energy cost. We will assume that 4 times that cost, or $4/kg is a practical lower limit. This is 400 times lower than the quoted price of the Falcon Heavy. Past studies have estimated that 98% of space systems could be made of local resources, and the remaining 2% would be rare or difficult to make, and thus more economical to transport from Earth. This gives a 50 times reduction in the amount of cargo that needs delivery. Combining the cost and traffic improvements gives a 20,000 times total potential improvement. That level of improvement will not be reached soon, and perhaps never, but the potential is so large it definitely justifies some engineering effort. How much is worth spending to get an improvement can be judged by the US $290 billion/year currently spent worldwide on space-related projects.

Developing New Systems Has Side Benefits:

- ***Technology Spin-off:***

Technology developed for one purpose or project often finds uses in other areas, a process called **Spin-off**, or technology transfer. One example is biochips for medical testing which use the technology developed for electronic microcircuits. Although there are many other such examples, they are unpredictable in individual cases. Space projects are not unique in generating spin-off technology, but they typically have a large amount of research and development effort, and thus more than average opportunity for it. There have been attempts to **quantify spinoff benefits** specifically from NASA programs. While we cannot put an exact value on spin-off technology, we can make an allowance for some amount as part of the reason to pursue new systems over current systems.

When space systems are intentionally developed with an eye to Earth applications it becomes less of a spin-off and more a matter of design intent. This is referred to as **dual use** or **multi-use** technology. For example, solar cells were first used on a significant scale to power satellites, but now there major application is on Earth.

- ***Expanded Markets:***

Significantly lowered cost to do space projects can expand existing markets for things like communications satellites, or open new markets such as asteroid mining for rare metals. Increased supply and lower cost for these metals can prompt new uses for them which are not economic at present. Therefore in addition to the direct benefits of lower cost, some estimate should be made for the secondary effect of expanded or new markets in the total value of a new system.

- ***Optimism for the Future:***

Fear of loss is twice as strong an emotion as opportunity for gain. If the Earth is seen as a finite, closed, zero-sum world then fear of the future can dominate. In that view people might take less risk and try to hold on to what they have. This perception is not reality - from an engineering point of view the Earth is open in energy and mass flows. However, perception can affect actions even if incorrect. An understandable path to opening the resources of the Solar System can change a pessimistic world-view to one with hope. There are other ways to change negative perceptions of the future, such as the development of affordable, clean energy sources. Opening space, though, directly refutes the idea of a closed Earth. A more optimistic world might be less fearful of new technology and more willing to invest in long term projects.

Continued on page 2

Section 5.1: Human Expansion (page 2)

back to page 1

Selection Criteria Approach

Having set the initial goals we want to achieve, we now want to start developing a design that best meets them, and evaluate how it compares to existing systems. One part of this is turning the vague term "best" into an objective measurement method. Complex systems such as these will have many parameters which can be used to compare alternatives. These parameters will have different units of measurement. Since different features of a design, like cost and performance, cannot be directly compared, we convert them to a unified numerical scale, and can then choose whichever option has a higher score on that scale. The conversion formulas represent a mathematical model of what is desirable to the **Customer**. In other words, it is an objective model of "best". This model will evolve through the conceptual design processAt this point we can only start to work on it.

Subjectivity of Criteria

The modeling and evaluation is an objective processThe choice of what parameters are to be measured and their relative importance, however, is subjective. It derives from human needs and desires which are not set objectively. In addition, humans often do not know what they would want because they are uninformed on a subject. In our present program, the ultimate customer is civilization as a whole. Most people do not have enough interest or information to know what parameters they would value. Therefore the immediate customer must act as a proxy who expresses their preferences for them. The immediate customers would be the ones in direct contact with the designers. For now that would be contributors to a research foundation, and the program designers themselves. At this stage they cannot educate and poll all of civilization to determine what they would want. Instead they must estimate what they would want if they were well informed and asked their opinion. As an example of this proxy method, the ultimate customers for a smartphone are the people who will eventually use them. Since you don't even know who those people will be, you cannot ask them what their preferences are. Therefore the management, marketing department, and engineering department act as a proxy for those customers and make their best effort to determine what theirpreferences would be.

Types of Criteria

Some criteria are a clear yes/no or pass/fail type. If the design fails to meet such a criterion, it has no value in the eyes of the customer. This type are not modeled or scored, since the score if not met would be zero. Instead they are included as well defined program requirements that must be met, and other parameters allowed to vary as needed to meet the "hard" (fixed and unchanging) requirements. Usually this type represents fundamental program goals and objectives. Other parameters, such as cost, have an incremental level of desirability - each increment of lower cost is more desirable. Such parameters can be scored on a sliding scale from desirable to undesirable.

Number of Criteria

Any parameter which has value to the end user/customer can potentially be used for design selection. Thus they can cover the full range of goals, requirements, and design features. In practice, some are more important than others, or only are relevant at certain levels of the design. A very large set of criteria is difficult to apply because you have to evaluate all of them for every design option. So the selection criteria are usually limited to the more important ones. Whichever criteria are chosen, there should be a clear description showing the relationship to customer desires and how the scoring formula is derived. We will give a couple of examples of cost criteria to show how this is doneAt this point they are not final selection parameters, merely examples.

Example of Cost Criteria

We have assumed that civilization as a whole is who would be paying for the program development as a whole and benefit from the results. Thus they are the ultimate customer from the Systems Engineering standpoint. One major reason for choosing a new design over current ones is cost, which is a measure of the resources input to the program. So we will use development and operating cost as our example criteria.In both cases, lower is better

- **Affordable Development**

Existing space programs are already quite substantial, indicating they are valued by society. Worldwide, government space agency budgets total US $33 billion as of 2011. The **Satellite Industry Association** estimates global space industry in 2011 of US $290 billion. Industry commonly sells to government programs, so to avoid double counting we only include 30% of government budgets as unique. This leads to a total of $300 billion/year for all space-related activity. Present value converts an annual flow of funds to an equivalent single amount. The current Price/Earnings ratio of the S&P 500 stock index is 12, implying a present value for current space programs of US $3.6 trillion.

A new or modified program should not greatly exceed what people apparently are willing to spend. Therefore we take the net development cost of a new combined system divided by the present value of current space programs as a measure of affordability. Small values are good, indicating not much extra cost. Negative values are even better, indicating cost savings. Some net cost is acceptable if the benefits exceed the cost. Net development cost here is the discounted present development cost of the system, including revenue generated. Thus it is the amount of money you would need today to finance all the future development at the maximum point, after accounting for any positive income that can cover later costs.

We do not have any absolute measure to convert net development cost to an evaluation score. In the absence of one, we will assume civilization is willing to spend the same fraction of total output on space programs. Since world real GDP grows about 3% annually, we will set an increase of net development cost based on the present value multiplier of 12 as setting an equivalent time horizon of 12 years. This gives 12x3% or 36%, as a nominal value, to which we assign a midpoint score of 50%. We will assign 0% net development cost a score of 75%, and 36% decrease in cost will score 100%. A 108% cost increase will score 0%, and the scale will extend beyond 0 and 100%. The score values are arbitrary. What matters is how they relate to the scores of the other selection criteria. This implicitly defines exchange ratios between criteria, like "a 1% increase in thrust is worth $5 million in development cost".

A better scoring of development cost could be obtained later by surveying people after explaining what the program benefits would be. We can also do a sensitivity analysis on the scoring. This is calculating how the results of the conceptual design change as a result of changing the selection criteria or their scoring. If you get the same design result for a wide range of scoring criteria, the design is said to be **Robust** or **Insensitive** to changes. This is desirable. If small changes cause very different design results it is likely not an optimal design, or more work needs to be done to distinguish alternatives or reduce uncertainty

- **Low Recurring Costs**

The previous measure reflects maximum net development cost at any point in time. A good conceptual design would also have a low ongoing operating cost.The major current recurring cost for space projects is launch to Earth orbit. Total launch cost is the product of how many kg you need to launch and cost/kg of the transport system. Improvements such as stronger materials or using space resources reduces the required kg for a given project. A number of component costs make up the recurring cost/kg, but the biggest improvements will come from diferent design and technology used for transport to orbit.

For this example we will measure cost reduction relative to current (2012) values as the baseline. Current space hardware and propellant masses for typical missions determine the total kg required. New designs and technology result in different masses for the missions. The mass improvement ratio is then current mass/new mass. For launch cost we will assume the quoted cost of US $1566/kg for the Falcon Heavy as the baseline. For new transport systems the marginal operating cost to Low Earth Orbit after paying for development is used.The launch cost ratio is then ($1566/kg)/(new launch cost)Total cost reduction is then the product of (mass improvement ratio) x (launch cost ratio). We will assume equal percentage reduction in cost has constant customer value, and therefore convert the logarithm of total cost reduction to a score. Half the maximum potential improvement, a ratio of 148:1, has a natural log of 5 and is scored at 100%.Other cost reductions are scored at 20% x ln(cost reduction).

Program Requirements Approach

The next part of developing the design is defining requirements. Program requirements will state how the general program goals will be met in terms of measurable features, parameters, and values. They will be documented at the end of the conceptual design stage. At this point (the start of conceptual design) we can start by identifying categories of requirements under each of the program goals. The requirements analysis step will then examine and select individual requirements from them.

Improving Life on Earth

This program goal was listed first because, for the time period we can reasonably plan a program, the majority of humans will live on Earth. The quality of our life here is therefore of high importance. Historically, and in the near future, almost all of the people working on space programs, the offices and factories, and sources of materials will also be on Earth. If new methods and technologies are developed to reach challenging space mission requirements, they can feed back to the toolbox civilization has in general to work with, and thus improve life here. An example of where this has already happened is in the Systems Engineering methods developed to manage complex aerospace projects. These methods can be applied to all forms of complex projects on Earth.

To relate this general goal to particular requirements, we can look at the impact of new space technologies on quality of life measures already in place. An example would be high efficiency closed loop food production. This could lead to improved food security or lower farm wastes when applied on Earth. Thus during conceptual design we should look for methods and technologies with Earth applications, determine what quality of life impact they could have, and give preference to those with the greater potential applications and impact. A fusion rocket would then be preferred over a regolith engine because functioning fusion reactors can also help solve energy security on Earth. A device that uses raw rock as reaction mass for propulsion does not have that kind of Earth application (that we know of).

Quality of life measures include physical measures such as GDP, life expectancy, and pollution levels, as well as social measures like education, leisure time, and civil rights. One step in establishing program requirements will then be to select the appropriate measures. The next step would be to set how much of our program should have potential impact on these measures, and what level of improvement is desired.

Understanding the Earth

Understanding our home planet is mainly a scientific enterprise. It includes gathering data, developing theories and models, and then testing those ideas against reality. After you have gained a better understanding, then deciding what to do with that knowledge is a social and political issue outside this program's scope. It may lead back to new projects such as orbital sunshades to reduce temperatures on Earth, but the choice to start such projects should be based on sound knowledge. This group of requirements will address gaining that knowledge.

We want to gather data about the Earth and other planetary systems. The reason for the latter is other planets are natural experiments in what happens under other circumstances. We cannot gain knowledge by experimenting with the one planet we live on, but we can test our theories and models by looking at other plants. Gathering data has dimensions of time and space. We want to know both the current properties of planets and their environment in three dimensions, and the history of those properties across time. We can then set program requirements in terms of increased detail in both. An example would be "Map all Solar System objects to a resolution of 1 kilometer or better." A large pile of data is not very useful by itself, so sufficient resources to organize it and develop theories is also needed, after which a new cycle of data collection to test the theories would happen. Program requirements would then include the scientific staff to make use of the data and guide later cycles of data collection, and presenting the results of the accumulated knowledge in a form that educates people on choices they may need to make.

Reducing Space Hazards

This class of program requirements includes identifying hazards from space, followed by preparing for or preventing the damage from them. Known hazards include asteroids and comets, solar flares, and stellar explosions. Types of requirements can include the ability to deflect hazardous objects, and limiting damage from events like flares that cannot be prevented.

Increasing Biosphere Security

A single uncontrolled biosphere is inherently insecure from natural and human-made variations. Since nearly all of us currently depend on this one biosphere, we want to set program requirements that improve the security level. The requirements groups of understanding the Earth and reducing space hazards help in reaching this goal. This group goes further to counteracting undesirable variations and adopting the idea of backups from computer technology. Specific requirement might include providing safe storage of biological samples and testing of biohazards away from the Earth, active control of biosphere parameters, and establishing artificial biospheres in space or on other planets.

Expanding Resources

All civilizations require resources to function. This group of program requirements can include identifying scarce physical, material, and energy resources and setting quantity and cost goals for increasing their availability. Physical resources include quality living, growing, and working space in terms of dimension and environment. Material resources are all the raw matter and specialty compounds and equipment needed. Energy resources are needed in various forms for different purposes. All of these are interlinked in terms of flows of resources for different tasks.

Long Term Survival

Requirements in this area would include examining long term resource depletion. For example, continued use of nuclear fission for power would eventually deplete the Earth of Uranium and Thorium. If that were a critical part of keeping civilization operating, then eventual collapse would occur. There are alternative sources of power, so this particular example is not fatal, but it shows the idea of looking for items that would run out in the long term. The Sun also gets 1% brighter per 100 million years from stellar evolution, and the Earth permanently loses about 3 kg/s of Hydrogen from water dissociation, equivalent to 850,000 tons/year of lost water. These long term changes will eventually make the Earth uninhabitable. So program requirements can be set to either counteract depletion and changes to the Earth, or enable moving elsewhere within the timescale of the problem becoming critical.

Increased Choice

Existence on Earth under our current civilization imposes natural and human-made restrictions. Measures in this area look at lifting or eliminating these restrictions. Some examples are freedom of location - on Earth you are restricted by national governments from living anywhere you want.Another is freedom of gravity - you cannot choose to live under a different gravity level right now

Increased Opportunity

Most of the Earth (the good parts at least) are already claimed by someone. By making new unclaimed or under-used areas accessible, that increases opportunities for people who want to start something new without first having to pay off previous owners. Measures in this area would include increased area or resources which are made available.

Design Approach

Having set up a way to measure how good the design is and establish requirement, we next need an approach to formulate the **System Concept**, a high level description of what the combined space system is and how it works. There is no single magic bullet (or magic rocket) that can meet all the program requirements by itself. If there were, someone would have used it by now, or at least be pursuing it seriously. Therefore we take the approach of ***leveraging multiple good ideas***, which allows the savings to multiply together. This will result in a complex program of multiple systems, which need to be combined for best results.

To meet the cost criteria of affordable development we do not build everything at once. Instead ***the ideas get applied in incremental projects and systems which build on each other***. This allows some return from the early parts to help pay for the later parts. The early parts are smaller in scale than what comes later, which further reduces initial development. This will result in a program which is extended in time.

Multiple Ideas

To develop our system concept we will use the following ideas:

- **Use less of or eliminate conventional rockets** - They have been in use for 50 years and had a lot of engineering development and optimization. Therefore using another conventional rocket is unlikely to bring much improvement, and other projects are already attempting to do so. Instead we will try to use some of the other hundred or so transport methods and variations identified in Part 2. This gives us the possibility of greatly improving on the performance and cost limits of chemical rockets which are imposed by their chemistry.

- **Design for re-use, repair, and recycling** - It should be evident that these features will reduce hardware and supply cost, and yet many launch vehicles and satellites are used once and disposed of. For human crews on, for example, the Space Station, oxygen and food supplies are similarly used once and disposed of. Some space hardware is designed for maintenance and repair, but much of it is not. Therefore we will try to incorporate these multiple use and long life features to get more use out of hardware and supplies.

- **Use the material and energy resources of space** - Again, it should be evident that bringing everything from Earth is a limiting factor, and the farther you go in space, the higher the cost of doing so. Solar power is so overwhelmingly useful that it has been used by almost every space project, but other material and energy resources have not yet been exploited. Therefore we will try to design for using them to leverage what we bring from Earth.

- **Build multipurpose facilities** - One-time missions, as have often been designed to date, tend to not leave anything useful behind. Thus the next mission is exactly as hard and expensive as the last one. Therefore we will try to design facilities that can be used multiple times or on a permanent basis. An example would be landing a solar array on the Moon that is used to recharge a rover vehicle quickly, and then later can also be used to power an extraction plant. This makes more sense than having a solar array attached to the rover, and then another solar array for the plant.

- **Use diverse modular designs** - Monolithic, or single piece, designs require replacing the entire item if your needs change or you have an upgraded technology you can use. For long term and complex projects you are not able to predict all the changes and upgrades that might be required. Therefore we will try to use modular designs where possible to make it easier to change things. Modular designs can also start smaller and be added to in steps, reducing initial costs and the size of transport systems. It is not required to use one method to do everything. On Earth we use industrial delivery systems like pipelines to deliver large quantities of goods cheaply, and reserve more expensive and safe methods to deliver people. Specialization of this sort is acceptable when it makes cost and technical sense.

Using the above ideas does not mean using them blindly where they are not appropriate. It means incorporating them where it makes sense, in an optimized amount. Past programs have tended to not use them enough, or at all. This has led to high cost and limited performance, which our new program tries to correct.

Incremental Projects

Rather than attempt to do everything at once, we will take the approach of designing and building our space program elements in progressive increments. The increments will establish new locations and improve measurable parameters in several dimensions as they get added:

- **Working environment range** - Starting with temperate Earth locations provide production, habitation, transport, and other system elements that work in the given environment. Then extend their working range to hotter, colder, wetter, drier, and higher and lower pressures. Later extend the range of working environments to space locations with the additional variables of gravity and radiation levels and increased range of temperature and pressure.

- **Time and energy range** - Time has components of communication time, travel time, and stay time. Energy has components of potential and kinetic energy to reach a given location. New locations will increase the range of communications and control distance, and require longer travel and stay times. They will also require greater energy changes to reach.

- **Performance levels** - These are measures like cargo capacity, industrial output, efficiency and closure, and how many humans are supported. They are designated as performance requirements for particular program elements and systems. Each location starts with a given performance set, and it is improved in stages to higher levels.

Increments which are far away in time or require parameters far beyond current experience become uncertain to design for. New technology may get developed in the interim times, and meeting untried parameter levels may be difficult. Therefore at some point it is no longer useful to plan future items in detail. Instead, the options for these items can be laid out, and a plan to develop needed technology and reduce uncertainties put in place. Therefore program engineering work is not done once and finished, but is a continuing effort. At any point in time the program will have a baseline documenting current status and future plans, but that baseline will get updated on a regular basis.

Initial Program Concept

Based on the above discussion we will describe a starting point for the program concept. We emphasize that this is only a starting point, the final concept is the endpoint of the conceptual design stage.

Program Description

Our basic concept for the program is to expand the range of human civilization to new locations while meeting the above program goals and requirements. Any kind of civilization seems to require the ability to produce food and other physical items, to provide shelter, and to move people and items from place to place. There may be other basic requirements, but we will start with these. Therefore for each new location we set up and then expand in stages the production, habitation, and transport elements. We start from existing civilization as it is, start a new location, and once it is sufficiently developed, travel to and repeat the process at the next one.

The first new locations will be easy ones on Earth, in temperate climates. New technology such as remote operations, automation, and resource extraction are first demonstrated there. Once built up, we transport new elements to more difficult locations on Earth, such as deserts, oceans, high altitude, cold, or underground. This expands the range of environments and the distances for remote control. Widening the range of conditions where people can live and work meets many of the program requirements, and the new locations should be self-supporting physically and economically

After sufficient locations on Earth are established, the production capacity is used to build transport to orbit, and the remote operations and other technologies are further developed for space locations and the more difficult environments of vacuum, temperature, radiation, and lack of gravity

To enable growth in new locations we design and build the following kinds of hardware:

Seed Factories

Developing a new location has always involved bringing a starter set of knowledge and tools. Historically that meant bringing animals, seeds, axes and hammers, and whatever else the technology of the time required to start building. For our future program, we want to use the best methods that modern technology allows. Our starter set should use automation and remote operation, and be able to not just make a fixed set of products, but make more equipment for itself to widen the range of outputs. So our concept is to use **Seed Factories** to establish production capacity at each new location. Unlike conventional factories which only produce a given set of products, a seed factory uses part of its production to make more equipment for itself. So over time it is able to make a wider range of products and use a wider range of local resources. The initial seed equipment, plus supplies and necessary components it cannot make yet, are delivered from the previous location. Over time, the production capacity will need fewer supplies and parts, and be able to make more items locally

Besides self-expansion, the factory output is partly items needed to live and work in the particular location. In space that would include items such as parts and supplies for habitats, vehicles, mining equipment, water, and oxygen. The remainder of the output would be commercial items to trade for needed outside supplies. On Earth that could be any kind of product with a market. In orbit, an example would be building large communications satellites. Vehicles would carry copies of the seed equipment on to the next location to start growth there. Since habitats for humans are part of the production output, at first you may not have the ability to support many of them at a new location.Therefore the seed factories are designed for a mix of automated/robotic, remote control, and direct human operation to minimize the latter. The first seed factories are built in various locations on Earth. When they have sufficiently expanded they use their industrial capacity to build launch vehicles and space hardware. This is used to establish assembly and processing equipment in Earth Orbit. This evolves into a full seed factory, and from that grows to other locations in space.

Cyclic Systems

A linear system might dig up a resource to be used as fertilizer, use it once, and then it gets removed in the harvested crops and runoff. In a crowded Earth, or in space locations, where getting new supplies is hard, resources for once-through linear designs become expensive and unsustainable. Therefore our concept includes using cyclic systems which take old items and send them back to earlier production stages for repair, re-use, recycling, or reprocessing to new items. Because transport is an overhead cost, we prefer to do the cycling locally but if it turns out to be more optimal to do the tasks at other locations we will do so.

High Efficiency Transport

New locations start with deliveries from previous locations, and will continue to need transport to deliver new equipment, supplies, and people. Once sufficient production is built up, delivery of finished products back to previous locations is needed. In addition, return of people and used items back to previous locations is needed before setting up permanent habitats and fully cycling systems. Thus transport is a necessary function, and we prefer to do that as efficiently as possible.

On Earth, machines like internal combustion engines are particularly inefficient, so we will look for more efficient replacements. In space transport, conventional rockets are even more inefficient reaching orbit than gasoline engines. Although chemical rockets are very efficient (~80%) as heat engines in themselves, when considered as part of a total transport vehicle, most of the work goes into accelerating fuel which is later burned and discarded. Therefore our concept is to replace conventional rockets as much as possible with a variety of higher efficiency transport methods. This includes several options for initial launch to orbit, and primarily electric thrusters once in orbit. Chemical rockets will not be entirely eliminated, especially at first, but are gradually replaced as the program evolves. Electric thrusters are about five to ten times more efficient in fuel use, and combined with the ability to extract fuel locally require dramatically less fuel from Earth, so we will try to use them heavily in the design. Like all rockets, chemical ones use fuel exponentially as a function of velocity. So even replacing part of their use will significantly reduce total launch mass. So our incremental approach still yields large gains even with smaller early steps.

continue to page 3

Section 5.1: Human Expansion (page 3)

back to page 2

We have set our starting point for the conceptual design stage by stating goals and potential benefits, initial selection criteria, classes of requirements, a design approach, and initial program concept. We now apply the **Systems Engineering** method from section 1.5 to further develop the design in increasing detail. The reader should bear in mind that a book such as this is a linear presentation of sections and pages, but the actual method is iterative, with the results of later steps feeding back to update earlier results. We will try to indicate when these differences between practice and description happen.

Program Requirements Analysis

Until now we have stated what the program is intended to accomplish in very general terms. The requirements analysis step takes these general statements and develops them into more specific and measurable features at the program level. In later stages of design these will be analyzed in more detail and assigned to separate systems and lower level items as **System and Design Requirements** which their designs attempt to meet. In the process we map the general requirements to the specific ones to make sure all of them will be met. The minimum number of necessary requirements should be defined at each level of a program's development. Excess requirements limit design options and impose extra costs to design for and verify

It should be noted that setting initial requirements does not mean they are physically possible, economically feasible, or a better choice than what exists now. They merely establish design targets, and later work will establish if they are the correct ones.

Requirements Sources

First we identify the sources or inputs to the requirements analysis. These should come from outside the program itself:

- **Program Goals and Benefits** - These were stated on page 1 of this section, and come from unmet desires of civilization as a whole, as best we can determine them.
- **Systems Engineering Experience** - Past engineering experience in developing complex programs has identified requirement types that often turn out to be necessary. We can use topical literature to identify ones relevant to this program.
- **External Constraints** - These are limits imposed by nature or from human causes such as legal requirements.

In addition to the external requirements, we have ones identified or derived internally from the program goals and benefits, design approach, or initial program concept, which are on page 2.

Analysis Process

Goals and Benefits - We examined each of the program goals and benefits as stated on page 1, and attempted to write individual statements of what the program will accomplish in terms of objectives, performance, schedule, cost, technical risk, safety, sustainability, and openness. The headings for these categories come from the list of requirement types in **Section 1.5**. Different programs would result in a different set of such statements. We tried to make the statements specific and have a measurable feature or parameter. In some cases, this is not possible yet and will have to wait until later in the design process. Placeholders are used for the incomplete statements. The result of this examination and a discussion of how each statement is derived becomes part of the initial draft of the Program Requirements in the next section.

Systems Engineering Experience- We next considered general categories of requirements, such as those listed in Section 1.5, to see what might apply to this program. The program is wide scope and long term at this level of analysis. Therefore many requirements that apply at more detailed levels are not relevant yet. One item that came up is the "Life Cycle", ie when would the program end? We included in the discussion under 1.2 that the endpoint is reaching the stated populations with the ability to continue supporting them afterwards. Compliance to laws, regulations, codes, and standards is too specific to apply at this level, as are durability and quality features. By it's nature this program is intended to have a positive impact on community and the environment. At more detailed levels these may become separate requirements. Manufacturing, Test, and Maintenance also apply at detailed levels. Flexibility, Scalability, and Evolution are expected to apply at the location or system level. Interoperation is met at the program level by virtue of the locations being a part of human civilization as a whole. At more detailed levels specific requirements may emerge for this category. The end result is not many new requirements were identified at the program level from this step.

External Constraints- For now, we consider the requirement to function in more difficult environments and distant locations as a sufficient identification of external constraints.For a specific location, such as the Lunar L2 point, the physical environment and legal regime will become inputs to the more detailed requirements.

Internal Program Requirements - Next we look at the classes of requirements identified on page 2 to see if any of these apply. From the "Improving Life on Earth" heading we define the requirement (2.4) for improved quality of life relative to current levels. The "Understanding the Earth" goal is indirectly enabled by inhabiting more difficult environments. When people are able to live in these remote locations, they can more easily study them, and science may be a major "occupation" for the residents. The data requirement (2.5) aids this goal explicitly. The category of "Reducing Space Hazards" is covered by the Population Risk (6.2) requirement, but we are not yet defining specific risks and reduction targets. This requirement also partly addresses moving hazardous research from the "Increasing Biosphere Security" heading. Active measures to increase biosphere security are covered under Survivability (7.2). Measures under the "Expanding Resources" category includes physical, which is dimensional in terms of locations, area, volume, and environmental range. This is covered by the primary program goal (1.1). Material and energy resources are covered by the Resources requirement (2.6). From "Long Term Survival" we added to the Survivability requirement (7.2) to design for depletion of critical resources. The "Increased Choice" heading is covered by the similarly named requirement (1.3). The primary goal (1.1) of widening the range of civilization addresses the "Increased Opportunity" category in the sense of access to unclaimed and under-used locations. The Openness requirements (8.) maintain access to the technology and locations for people outside the program, giving them increased opportunity.

Design Approach - Our next step was to look at our design approach to see if that drives any program requirements. Normally design progresses from requirements to solution, but this is a check to see if our approach feeds back in the other direction. The idea to reduce conventional rocket use is not a program level requirement, it is a design solution for the space part of the program. The program requirement to lower Earth launch cost (4.3) is what will likely drive the use of that idea. This requirement will also drive using reuse/repair/recycling in terms of longer vehicle life. Improved technology (2.3) and surplus resources (2.6) along with other requirements also drive reuse/repair/recycling, as that idea is a general solution to design optimization. For the space part of our program, using the resources of space is another general solution to optimizing the design. The resource requirement (2.6) to have a surplus output will drive using this idea. Building multipurpose facilities is another optimization vs single purpose design for each task. The program scale requirements (1.2) for permanence and size will likely drive using this idea. Along with growth (2.2) it will drive using the idea of modular design. The design approach of incrementally increasing environment range, distance, and performance has already been incorporated into multiple requirements. Thus we find the requirements will tend to use the ideas in our design approach, but the ideas do not force new requirements.

Program Concept- Similarly we check our initial program concept from page 2 to see if any new requirements are identified:

- **Program Description:** We identified a need for production, habitation, and transport as basic to civilization, but these will be developed as lower level functions for each location rather than top level requirements for the program. The top level is covered by the requirement to support a certain number of people (1.2). Set up and expansion of these functions is covered by (2.2) Growth. The main program goal (1.1) is to move to more difficult and remote locations. This will drive remote operations and automation, because humans can't live there until habitats are in place. Resource extraction is explicitly included as a requirement (2.6). Specific Earth or Space location are not identified as program requirements, but are implicit in number of locations (2.1).

- **Seed Factories:** The idea of using factory output to expand itself is a good solution to address the Growth (2.2) and Improved Technology (2.3) requirements, but does not impose new requirements at the program level. Technical details of such factories and what they produce would be lower level design and requirements.
- **Cyclic Systems:** The idea of using cyclic flows as much as possible is also a good solution to meet many of the program requirements, such as 7. Sustainability and 4. Cost, but we do not see new top level requirements from this idea.
- **High Efficiency Transport:** Again, this is a good idea for cost and efficiency reasons. We have set an Earth Launch Cost requirement (4.3) which will direct how hard we work to include this idea. The challenging cost target will tend to force reduction of chemical rocket use as a solution, but this does not impose new top level requirements.

For now, this completes our analysis process. As with any engineering design, more detailed work later on may cause an update to these results.

Initial Draft of Program Requirements

The following requirements list is a first draft drawn from the process described in the previous section. The individual requirements are sorted by category and numbered for ease of tracking, with a discussion of how they were developed. Further conceptual design work will add to and refine this list to produce a final **Program Requirements Document**, one of the outputs of the Conceptual Design stage of a program.

1. Objectives

- **1.1 Program Goal** - The program shall expand human civilization to a series of new locations with increasingly difficult environments and distance.

Discussion: Our program concept is to develop new locations starting with the easiest, but with improved technologies in areas like energy, automation, and recycling. This should result in lower costs and improved quality of life measures. This first requirement is just restating the first part of the concept as a definite objective. It is quantified and refined by all the following requirements. Note that new locations include dificult ones on Earth as well as in space.

- **1.2 Program Scale** - Expansion shall be demonstrated by permanently supporting at least 95,000 humans total among new Earth locations and at least 2,000 humans per new space location.

Discussion: A target size is needed to know when you have reached the program goal, and to determine the size of hardware designs. They are intended to be large enough to prove the new locations are permanently occupied and the new technologies function reliably. They are end-point targets for this program, but nothing would prevent further expansion afterwards, or in parallel by other programs. Life cycle analysis ends at this point, but the locations are designed to continue operating. The initial numbers of people will be much smaller, possibly zero if robotic and remote operation is the best way to begin building in space. As a starting point the numbers were arbitrarily chosen to be the square- and cube-root of total human population in 2050. Later sizing analyses will likely change them, but something is needed for the first round of design work.

Demonstration of the new locations and meeting all the other requirements supports the benefit of optimism for the future. It provides an understandable path to an growing, open, and improved life, rather than a finite, closed, and zero-sum world.

- **1.3 Choice** - Specific locations and their internal organization, function, and operation shall be chosen by program participants and location residents within the limits of design constraints.

Discussion: This requirement is included to implement the goal of increased choice and freedom. Therefore choices such as where to build a new location are not imposed from above, but made internally. Participants are people like the designers, who may not end up living in the new location, and residents are the ones that do. Constraints are limits on a design imposed by nature, like level of sunlight at a location, or by humans, such as a law which prevents using fission reactors for power at a given location.

2. Performance

- **2.1 Number of Locations**- The design shall maximize the number of new locations, where new is defined by at least a 10% increase in an environment parameter or distance measured in time or energy terms.

Discussion: Requirements can be a fixed number or a variable parameter. Here it is a variable denoted by "maximize" without a specific number. A requirement which can increase without limit results in an unbalanced or infeasible design. So desirable parameters like "more locations" are balanced against the others by using measures of effectiveness and scoring across the entire program. The second part of the requirement establishes how much difference is enough to be new in terms of location.

- **2.2 Growth** - Each location shall increase the capacity for production, habitation, and transport in a progressive manner.

Discussion: Our initial program concept identifies these as basic functions of civilization. Setting a requirement to grow rather than meet a given level all at once allows incremental and modular design. The growth rate and target capacities for each function are not specified at this point. Later modeling and optimization will determine what is feasible, and specific values will then be applied to selected locations.

- **2.3 Improved Technology** - Locations shall increase the levels of self-production, cyclic flows, and autonomy in a progressive manner.

Discussion: The program concept is not just to expand to new locations, but to improve technology levels. This requirement sets the main categories of improvement. It is in terms of all locations as a group, since one location may be dedicated to a single task like mining or growing food. Different locations then trade outputs. The choice of what to do at each location and how much transport is needed between them is a design optimization to be settled later. Self-production is the amount done internally rather than obtained from outside sources. Cyclic flows are materials which get recycled and reprocessed, rather than new inputs and waste outputs. Cyclic flows do not count new production for internal growth or external delivery. Autonomy measures levels of automation and internal control so that a location does not require as much outside support or human labor to keep operating. The numeric levels for these parameters will be set later by what is feasible.

- **2.4 Improved Quality of Life**- Completed locations shall provide an improved physical and social quality of life relative to the upper 10% of Earth civilization.

Discussion: This makes specific the goal of improving life on Earth by demonstrating quality of life well above the current average. The exact measures to compare to will need to be defined later in the concept development. The 10% threshold is a notional goal, representing "better than most" conditions.

- **2.5 Data** - The program shall collect and disseminate [TBD] data about the Earth's environment, surrounding space, and objects therein.

Discussion: This requirement supports the program goal of understanding the Earth better by looking at its environment, and the history and evolution of other planets and objects. We don't know at this stage of the analysis what kinds of data will be useful, at what level of detail, or at what cost, so we use a placeholder value. Existing astronomy and planetary science program are already doing this task, so the final requirements will be in terms of what capabilities beyond that the program will add. In addition, some of the data collection is needed to support future expansion to new locations in space.

- **2.6 Resources** - The program shall output a life cycle surplus of at least 100% of internal material and energy resource needs.

Discussion: - This supports the program goal of expanding the material and energy resources of civilization. The program will consume resources internally to operate, and will not produce a surplus immediately, so this requirement sets a net output requirement for the total life of the program back to civilization.

3. Time

- **3.1 Completion Time** - The expansion to a new location shall be completed before expected progress in technology indicates a re-design is required.

Discussion: If the program is too slow to develop a new location, then re-design to account for new technology would be indicated before you are finished, and therefore previous design work is wasted. This is separate from economic or other reasons which might indicate a particular schedule.In case of multiple requirements driving one parameter, the strictest one is used.

- **3.2 Operating Life**- Locations shall be designed for an indefinite life, with maintenance, repair, and replacement.

Discussion: Expansion of civilization is an open ended process, so the intent is to set up permanent new locations. Anything made by humans eventually wears out, breaks, or could be replaced by something better. To reach an indefinite life, then we need to provide for ongoing maintenance, repair of failed items, and eventual replacement or upgrade.

4. Cost

- **4.1 Total Development Cost**- The total development cost for new technology and hardware designs shall be less than 50 times the unit cost on Earth, and 5 times the unit cost in space of the hardware.

Discussion: There are only a small number of governments and businesses capable and motivated to do large space projects. A reduced scale, especially in the earlier stages, means more possible entities that could perform the projects, and thus more chances it actually gets done. Large entities are not excluded from these kind of projects, they just should not the only ones able to do them. This is one reason to limit the total development cost. Another is to generate an economic rate of return so the program can pay for itself (or at least pay part of its way). We set reasonable values of 50 and 5 times the unit cost of hardware for Earth and space based on an estimate of the number of locations the hardware will be used, and the desire to keep the development cost reasonable in relation to unit production cost.

- **4.2 New Location Cost**- The peak net project cost for a new location shall be less than 50% of the expected long term net output.

Discussion: For a new program to be justified over doing nothing or continuing with existing programs it must show some benefit or improvement. This requirement sets a minimum benefit in purely economic terms that the life cycle output be worth twice the peak project cost. There are other types of benefits than monetary ones, which would be covered by other requirements, and more detailed cost goals will be worked out in later stages of design.

- **4.3 Earth Launch Cost**- The program shall progressively lower the Earth launch cost component of total system cost, with a goal of $0.08/kg of total system mass.

Discussion: One of the potential program benefits is dramatically lower costs for space projects. We include this as a requirement in order to drive the design to explicitly attempt to reach it. This only addresses the space transport part of the program, Earth surface transport for terrestrial locations is already well optimized, and any program requirement to lower it would be separate and a much smaller goal. Assuming local resources are used in space, a reasonable long term goal is for 2% of total mass to come from Earth. Therefore the goal for the launched mass would $4/kg. The product of the two factors results in Earth launch cost, so the component requirements are inversely related. This is a long term goal, and will not be reached in early parts of the program. Total system mass includes consumed mass like propellants, so propellant production may reach very low percentage launched from Earth fairly easily.

The combination of improved technology (2.3), surplus resources (2.5) and this requirement supports the program benefit of expanded markets.

5. Technical Risk

- **5.1 Risk Allowances**- Program designs shall include allowances for uncertainties and unknowns in knowledge, performance, failure rates, and other technical parameters. New designs with higher risk can be included in program plans, but a process shall be included to resolve the risk, and an alternate design with lower risk maintained until resolved.

Discussion: Note that technical risk comes from design unknowns, and is distinct from accident risk. The program goal is to develop new locations with more difficult environments using new technology. The nature of pushing boundaries like this means that a reasonably large level of technical risk will be encountered at first. This does not mean taking risk for it's own sake - there should be sufficient potential gain to justify the higher risk. We adopt an allowance method to manage the technical risk, meaning we estimate the uncertainties and include them in our design. Rather than passively carry the risk, we have added requirements to reduce the risk by methods like testing, and carry a design alternative as a backup if the new technology turns out not to work as desired. The alternative to managing risk it to be safe and conservative, but that does not lead to progress.

6. Safety

- **6.1 New Location Risk** - New locations shall progressively lower internal risks to life and property, with a goal of significantly lower risk than the general population.

Discussion: Expanding to new and difficult environments will inherently involve some risk. It is acceptable to have higher than average risk at first. Many people willingly accept such risks if they know about them. This requirement sets a goal that once a location is fully developed, it should be relatively safe for the people and property internal to the location.

- **6.2 Population Risk** - The program shall significantly reduce natural and human-made risks to the general population, including external risks created by the program.

Discussion: There already exist natural risks such as solar flares and asteroid impacts, and human-made risks from orbital debris and launch accidents. These are hazards to society in general. To meet the program goal of reducing hazards from space we set this requirement to reduce them. The program itself will create some new hazards and risks to people and places outside itself. The goal is to reduce the total, including the newly created ones.

7. Sustainability

- **7.1 Biosphere Security** - The program shall increase biosphere security by establishing alternate biospheres and long term storage of biological materials.

Discussion: All life on Earth and human civilization currently depends on the single natural biosphere. This is a potential single point of failure from natural or human-made causes. Having a backup in the form of additional functioning biospheres and storage increases security. The habitation requirements under Performance could be met entirely mechanically, so this requirement adds biological features. We cannot reasonably duplicate the entire Earth, so the scope of this requirement will be set by what is feasible.

- **7.2 Survivability** - The program shall design for the long term survival of life and humanity from changes to the Earth which will render it uninhabitable and depletion of critical resources.

Discussion: It is well known that the Sun will get hotter and eventually become a red giant. There may be other long term changes, such as to the Earth's orbit. This requirement sets in place design features that will enable reacting to those long term changes. Because they are far in the future, it may not require much action in the near term beyond understanding what the nature of those changes will be. Critical resources are ones necessary for civilization to function and have limited sources and will run out at some point. Part of the design is to consider what these are and look for alternatives or additional sources.

8. Openness

- **8.1 Open Design** - Technology and design methods developed within the program shall be open for others to use. Specific instances of a design and produced items may be proprietary

Discussion: The combination of developing Earth locations first (1.2 and 2.1 above), improved technology (2.3) and this requirement are aimed at the benefit of technology spin-off/Earth applications. Since they are used on Earth first, and made open, there should be substantial transfer outside the program.

- **8.2 Access** - Development of a new location shall not prevent reasonable access for transit or to unused resources.

Discussion: 8.1 and 8.2 both address the goal of increased opportunity. Opening general technology and methods allows others outside the program to make use of them. Specific designs and artifacts can be proprietary because of the necessity to cover costs. Barriers to entry reduce opportunity, so allowing later arrivals to travel through or access unused resources minimizes those barriers. An example of unreasonable claims would be an exclusion zone extending 1000km from a high orbit platform. Some exclusion zone is reasonable for safety and to prevent sun shading, but not that large.

contine to page 4

Section 5.1: Human Expansion (page 4)

back to page 3

Program Evaluation Criteria

As noted on page 2, we want an objective method to choose among design alternatives, and to evaluate this program or its parts against existing programs. For a complex program such as this we take multiple measurable features or parameters and convert them to a common scale. The features chosen are the ones of value to the ultimate customer, which here is civilization as a whole, and the conversion formulas are according to their relative importance and desirability. Since we cannot ask everyone in the world what they want, we have to act as a proxy for them and make our best estimates of what they would want if they were well informed on this topic. We can draw on outside information to help with this process. Each design alternative will have different actual feature values, and result in different scores when converted to the common scale. Evaluation of the alternatives then amounts to adding all the scores and seeing which one has the highest total. For this step of the Conceptual Design we can only establish what our criteria should be. Devising alternatives and selecting among them will come later

Identifying Candidate Criteria

Every possible measurable feature or specification could be used as a point of comparison, but this is unfeasible for two reasons: (1) the time and complexity to evaluate all of them for every alternative, and (2) many features are simply undefined at a high level of a program. So we restrict our candidates to those which are the most important, and those which can at least be estimated at a high level. Any feature which is absolutely required for the program is not a point of comparison, since all the valid alternatives must include it. What that leaves us with are parameters which are variable in some way, so that an alternative can do better or worse by that measure. An example of a good measure is cost. Every alternative has an associated cost, which is almost infinitely variable, and so useful for comparisons.Most people will agree that lower cost is better, although they may differ by how much better. The general agreement that cost is important and which direction is better allows setting up a formula to convert particular cost ranges to score values.

We already (page 3) developed our program requirements based on program goals. By designating them as requirements we have indicated they are important, so those are the first places we should look for measurable criteria. Beyond that, we will also look at the same headings we used for the requirements analysis process, and consider civilization needs and desires generally, as we can find them from outside sources.

Program Requirements

Referring to the **Program Requirements** on page 3, we identify our first set of candidate criteria as follows:

1.0 Objectives:

- **1.1 Program Goal** - This sets the fundamental goal of the program, which must be met. With no room for variation it provides no candidate criteria.
- **1.2 Program Scale** - This sets a requirement of number of people on Earth and in space to be supported. The intent is to demonstrate permanent occupation and use of a location, but this can be done with a different number of people than specified. This makes a good candidate.
- **1.3 Choice** - At this level it merely requires choice by participants and residents, but does not specify how much. Since we could define degrees of choice, this can be a candidate at this level or lower levels. Criteria applied only at lower levels are either an absolute requirement at the program level, or summed somehow to a higher level result.

2.0 Performance:

- **2.1 Number of Locations** - The requirement is to "maximize", but does not state a specific number. This is an excellent candidate because it has an inherent range. An unbounded number of locations would tend to result in an unbounded cost, so this criterion would need to be balanced against cost criteria.
- **2.2 Growth** - The requirement is to increase capacity progressively, but without a number specified. This is another good evaluation criterion candidate. We can define measures for absolute capacity, increment size, and growth rates.
- **2.3 Improved Technology** - Like growth, technology levels are required to increase progressively, but without specified values. We can define ranges for this parameter and weigh it against other criteria.
- **2.4 Improved Quality of Life** - This sets a lower threshold of top 10% of Earth civilization, but no upper bound. The specific physical and social measures will make good criteria.
- **2.5 Data** - The goal is to collect and disseminate an unknown amount of data. This is inherently a variable parameter, so we will include it as a candidate.
- **2.6 Resources** - We have a requirement of 100% surplus of resource needs. Since a design might fall short or exceed this level, it is a good candidate.

As a note, performance levels of a complex program are typically a rich area for finding evaluation criteria, since they tend to be variable with design choices.

3.0 Schedule:

- **3.1 Completion Time** - The goal here is to complete a location before technology renders it obsolete and a redesign or replacement is indicated. We can analyze this and set a specific time range for completion as a criterion.

4.0 Cost:

- **4.1 Total Development Cost** - Like performance, cost is another rich area for evaluation criteria, since cost is a barrier in the physics and social sense to a program getting approval. The ratio of performance to cost is a particularly popular measure of effectiveness. This is a good candidate.
- **4.2 New Location Cost** - This sets the net project cost less than half the long term net output. This is an explicit effectiveness measure, and makes a particularly good candidate criterion.
- **4.3 Earth Launch Cost** - This is another progressive improvement goal with a variable value and an aggressive target, so it makes a good candidate criterion. At lower levels it can be divided into component measures for launch cost/kg and percentage mass from Earth.

5.0 Technical Risk

- **5.1 Risk Allowances** - The requirement is to include allowances for technical risk in the design. Since the margins for unknown design factors is already included in performance and cost estimates, it would be double-counting to also evaluate those margins directly as a criterion. A smaller weighted criterion would be reasonable for the size of the margin, as less uncertainty in the design would be considered better than more uncertainty

6.0 Safety

- **6.1 New Location Risk** - This is internal risk to the program contents and people. As a variable parameter, it is a good selection criterion.
- **6.2 Population Risk** - This is external risks outside the program including natural hazards already in place. Again, it is a variable parameter, and thus a good candidate.

7.0 Sustainability

- **7.1 Biosphere Security** - Security in and of itself is not a measurable parameter, but diversity of locations and species in alternate locations is a candidate.
- **7.2 Survivability** - This is a requirement for long term survival. A measure would be the rate of compensation vs expected time for a problem to become critical.

8.0 Openness

- **8.1 Open Design** - The program will either be open or not, so this is not a good candidate. We will assume that all design alternatives will be open.
- **8.2 Access** - This is another fixed requirement for reasonable access, which will be included in any design option, therefore not an evaluation measure.

This gives us a good starting list of candidate criteria. In early conceptual design many of the values will be undetermined for a given option. When not enough information is available to make a clear choice, the proper course is to keep multiple options until you have enough information to decide.

Additional Candidate Sources

- **Program Goals and Benefits** - Looking back at these on page 1, we find they are well represented by the requirements and don't immediately present new candidate criteria.
- **Systems Engineering Experience** - Reviewing the Systems Engineering section (1.5 of this book) does not present new program level criteria. Many of the requirements categories will apply at lower levels.
- **External Constraints** - Limits imposed by the natural environment or human rules are not variable parameters, they must be met. Therefore they are not candidates to compare options.
- **Internal Requirements Classes** - We review the paragraphs on page 2, and note that "Improved Quality of Life" needs more specific measures. For "Understanding the Earth" we note that data requirements should trace to useful results which can be acted upon, though how to measure that is problematic. Under "Biosphere Security", ability to counteract changes can be a variable measure. Under "Expanding Resources" we can measure the increase in economically viable available resources, including physical space.
- **Design Approach** - From the list on page 2, we can identify percentage of: closed cycle, local resources, self production, and reduced human and remote control inputs as variable components of the improved technology criterion (2.3). For number of locations (2.1) we can list temperature, rainfall, pressure, gravity, and radiation levels as environment parameters to expand upon. Time parameters include communications, travel, and stay times. Energy parameters include potential of location, and available flux from natural sources. Another possible measure is how many parameters and by what amount is their range increased per location, or maximum total increase. For all these criteria, existing state of civilization is the baseline level.
- **Program Concept** - This does not appear to have new criteria.

General Needs and Desires

NOTE: Improvement Needed This section is preliminary, and an opportunity for improvement.

Here are some ideas about general human/civilization needs and desires. The idea is that since civilization as a whole is the "customer" for this program, it is their needs and desires for which we should be designing, and therefore include these types of items in the evaluation criteria. The following items are drawn from online search, and are not yet backed rigorously or empirically. Until better defined, we are not using them as evaluation criteria.

- **Maslow's Hierarchy of Needs** - This is the psychological theory originated in 1954 that people address more basic needs before higher level ones. The general levels are (1) Physiological, (2) Safety, (3) Love/Belonging, (4) Esteem, and (5) Self-Actualization. This theory has been criticized. A recent paper by **Tay and Diener** may shed some empirical light on this topic.
- **States of Being** - The urge to be part of something larger than oneself, engagement with an item or activity, improved personal productivity or life, positive being or enjoyment, personal well-being/meaning/fulfillment.
- **Categories of needs and desires** - Mental/intellectual, emotional, and physical: see **Applied Empathy**
- **List of emotional needs** - The need for, or need to be: accepted, accepting, accomplished, acknowledged, admired, alive, amused, appreciated, appreciative, approved of, attention, capable, challenged, clear (not confused), competent, confident, developed, educated, empowered, focused, forgiven, forgiving, free, fulfilled, grown or growing, happy, heard, helped, helpful, important, in control, included, independent, interested, knowledgeable, listened to, loved, needed, noticed, open, optimistic, privacy, productive, protected, proud, reassured, recognized, relaxed, respected, safe, satisfied, secure, significant, successful, supported, treated fairly, understanding, understood, useful, valued, and worthy. Everyone feels these needs in different amounts.
- Look at empirical behavior to find out what people really want, vs what they say they want when interviewed. For example, working hours as income goes up between countries, over historical time, or within economic groups could be extrapolated to find out what people would do if work was not required.

Selecting and Weighting Candidates

Now that we have established a set of candidates, preliminary as it is, the next step is to choose the most important ones, and establish relative scoring weights and conversion formulas. The weight is how much a given criterion contributes to the total score of a given design. The more important that feature or parameter is, the more weight we give it. The conversion formula takes the feature or parameter and converts it to a relative score, usually with a nominal 0 to 100% range, or 0.00 to 1.00 value. The range is arbitrary, as long as it is consistent within a project. We will use the 0 to 100% range. The total score is obtained from the weight x score for each component criterion, then summing all the resulting products.

We want to narrow the list to the most important ones because it takes too much time and effort to evaluate many criteria for every design option. One way to do this is by making some of them fixed requirements, which all options must meet. Another is to simply drop the candidate as not important enough at the top program level. Whatever list of criteria is developed through this process, they should be reviewed by the rest of the program participants and customer as best is possible. Setting criteria and importance is inherently subjective, so it is very important to get agreement and consensus from the right group of subjective humans that these are the right set to use in evaluating the design.

Candidate Discussion

The following discussion gives our reasoning for the conversion from program parameter or feature to a score. We use the same numbering as for the requirements to make them easier to compare. Since not every requirement has corresponding evaluation criteria, there are gaps in the numbering

- **1.2 Program Scale** - Studies from animal populations indicate below 100 members, long term viability is threatened. Human social research gives **Dunbar's Number** for cohesive groups, ranging from 30 up to about 2500 for various types of groups. We therefore set 100 people per location as a minimum goal, gaining a score of 0%, with each factor of e above that gaining 25% in score. Thus 5460 people per location yields a 100% score. Similarly, we set the total for all locations on Earth at 5000 for a score of 0%, and 273,000 to 100%. We have a preference for a diversity of locations rather than one big one, since the overall goal is to expand to a series of multiple more difficult ones. Therefore we will weight the total population goal 1.5 times as heavily as the per location goal.

- **2.1 Number of locations** - Humans already exist in a range of environments. We will define the range where 90% of the population already lives as the baseline, and increments of at least 10% beyond that in at least one parameter a "new". Although 5% of the population already lives at each extreme, when combined with other requirements for improved technology, it leads to new designs Parameters which can fall to zero, like pressure or temperature have linear increments down to zero. With unbounded values the 10% increments are compounded ratios. Note that environment parameters are for the external environment, not the internal living and working conditions. The following parameters have been identified so far:

Environment:

Environment temperature - range of daily high and low across seasons in Kelvin (K).
Water supply - annual rainfall + running water/ice/air moisture flow in meters.
Atmosphere pressure - average value at the location, in kiloPascals (kPa).
Ground pressure - foundation design load in megaPascals (MPa) or exterior water or rock pressure for below-surface construction.
Energy supply - Flux from natural sources in W/m^2.
Gravity level - only applies to space, in meters/sec^2.
Radiation dose - measured by human biological effect in Sievert/year.

Time and Distance: These are measured from 5th %ile nearest population on Earth.

Ping Time - minimum round trip communication delay, in seconds.
Travel Time - one way normal travel time for humans.
Stay time - average per person stay time per location in years. Increments count linearly from zero.
Transport energy - total potential, kinetic, and frictional energy to reach the location for most efficient cargo method, in megaJoules/kg.

The two criteria we derive are the actual number of distinct new locations the program establishes, and the total number of range steps expanded to by all the parameters combined. The first is measured directly at 1% per location, and the second at 0.5% per range step.

- **2.2 Growth** - The fastest growing national GDPs are about 10% per year, so we will give that a score of 75% here. Average GDP growth is about 4%, so we will adopt 5% as a score of 25%. We will use the average of industrial output, resident population, and transport capacity to measure "output" of a location, and growth rates relative to the final size of the location. So the minimum and maximum rates of 2.5 and 12.5% imply 40 and 8 years respectively to grow to final size, and sets minimum initial size as 1/40 of final size.

- **2.3 Improved Technology** - These are measured at the program level across all locations. The percentage of local resource use, self production, and cyclic mass flow all scale directly from 0 to 100%. Automation is measured in reduction of human labor hours per output relative to current technology. Autonomy is the percentage of internal human labor and control at the locations relative to the total required for the locations to function. These are both scored directly in percent.

- **2.4 Quality of Life** - We will set a nominal GDP per person of $50,000 in 2012 as the 25% score level. This is about the average for the top 10% of world population. There are many other potential quality of life criteria, but for simplicity we will use just this one for now. The highest single country GDP is Monaco at about $180,000, so we will set that to 100%, and scale linearly

- **2.5 Data** - At this point of the conceptual design we do not know enough about data to set it as a criterion, therefore we put it aside for now

- **2.6 Resources** - Our nominal requirement is 100% surplus, and more is better. No surplus we assign a score of 0%, since it is not meeting the desire to improve life outside itself. We give each doubling of output a 25% increase in score, thus 200% output (100% surplus) scores 25% and 15 times surplus (16 times output) scores 100%. We are choosing a local measure of resources relative to the program rather than global availability. If civilization as a whole wants to increase its resources, it can copy the programs examples.

- **3.1 Completion Time -** This seems to duplicate 2.2 Growth as far as setting an overall time to reach a final size. For the present we will not list it as a separate criterion.

- **4.1 Total Development Cost** - For terrestrial locations we can set one-time (non-recurring) development cost in the range of 10 to 100 times the unit cost, on the principle that multiple copies of locations will eventually be built, and the one-time cost will be distributed. For space locations we expect fewer copies, and that some key technologies will have been previously developed for Earth. Thus a range of 1 to 10 times unit cost is more reasonable. Since lower cost is better, we invert the development/unit cost ratio and multiply by 1000 and 100% respectively to get a score scaled to 100%. For difficulty of the location beyond temperate, we allow 10% new development cost for any environment parameter step above those previously used. For resident capacity we scale by ln(actual/nominal) size of 75 people, to account for larger or smaller elements.

- **4.2 New Location Cost** - This is the explicit unit cost per new location relative to total output. As written, the requirement overlaps with 2.6 Resources, so we instead use an absolute cost per person, with the US total capital per person set to 50% score. Each factor of 2 up or down adjusts the score by 25%. For space locations we allow twice as much capital cost per person. We allow for the difficulty of the location by adding 10% linearly for each environment parameter step beyond the temperate range.

- **4.3 Earth Launch Cost** - This cost is the transport to LEO component for space locations in $/kg of total system mass. This includes mass obtained locally in space. Because both actual transport to orbit and use of local mass (currently 0%) can be greatly improved, we use a steep scoring function. A current baseline of $1600/kg gets a score of 0%, and each factor of 10 reduction will score 20%. Thus the requirement goal of $0.08/kg will score 86%.

- **5.1 Risk Allowance** - Less variation is better, so we score this on an inverted scale. A technical risk margin due to design uncertainty of 50% would score zero, and a margin of 0% (which is only reached with completed and tested designs) would score 100%. More advanced technology may give better potential performance, but with more uncertainty. This can be reduced by development and testing, but that is in the future. Present uncertainty is used to evaluate this criterion.

- **6.1 New Location Risk** - The goal is significantly lower internal risk, although higher risks are acceptable as a temporary measure while setting up. We scale this as equal to current general population risk gets a 50% score, and each doubling of risk lowers the score by 25%, each halving of risk increases the score by 25%.

- **6.2 Population Risk** - These are risks to the external general population from program or natural causes. Because the whole world is affected, it is difficult for one program to have much effect, so we give this a narrow scoring function. The program acts more as a demuonstration that the risk reduction is possible, and civilization can exert itself if desired to do more. Each 5% reduction to existing population risk is worth 25% score. We base no change to population risk at 0% score since any increase in total risk is generally considered unacceptable for a new program.

- **7.1 Biosphere Security** - Maintaining biospheres outside their natural environmental range increases security by having backups and the ability to survive transient disruptions. Zoo breeding populations of endangered species and seed banks are examples of existing programs of this type. It is difficult to say how much of this activity is enough, so we will somewhat arbitrarily score total number of species x locations. For each factor of 10 increase starting with 10 we will add 20% to the score. Thus 100,000 species in 10 locations is 1 million total species-locations, and would score 100%.

- **7.2 Survivability** - Like population risk, a single program cannot guarantee long term survival by itself, so we set a narrow scoring function. For each 5% compensation for long term change and resource depletion the program reaches during its life it gets a 25% score. For a change like the Earth overheating due to the Sun, which might take millions of years, only the change which occurs during the program duration (perhaps 50 years) is being compensated for. For critical resources, only those without which civilization cannot function are considered. Ones with reasonable alternatives are not critical. Compensation can be by active measures, like shading the Earth from overheating, or by alternatives like moving to other planets.

Weighting Discussion

Next we discuss our reasoning for the relative weights of the criteria. We will use a total weight of 100 points for all the criteria together. Our weighting is subjective, based on human opinion as to the importance of design features and parameters. Design alternatives themselves are objective. So other people reviewing the program choices can simply change the scoring and weighting to fit their own opinions on what matters. The design alternatives can remain the same, but a different set of choices would result from the changed evaluation scores.

Most people, all other factors being equal, prefer to get more results relative to the cost or effort expended. Since the ratio of performance to cost can be equally affected by increasing performance or lowering cost, the relative weight of these criteria groups is often set to be equal, and a large part of the total weight. In our list above, the performance type criteria are from 1.2 through 2.6, and the cost type criteria are 4.1 through 4.3. The remaining criteria fall into the technical risk, safety, and sustainability categories. Historically, large complex programs tended to focus on cost and performance, and put relatively small weight on other factors. With feelings of a finite and connected Earth, longer lives, and greater wealth and standards of living, people feel there is more to lose and so place more importance on the possible negatives of a project. We expect this trend to continue in the future, and this will be a long term program, so we will assign 30 points to this group, and divide the other 70 points equally between performance and cost.

Performance Group (35 points)

Scale (1.2) and number of locations (2.1) are the main motivations for a program of human expansion, so we will assign each 7.5 points. There does not seem to be strong reason to give growth, improved technology, quality of life, or resources (2.2-2.4, 2.6) more or less weight among themselves, so we assign 5 points to each.

Cost Group (35 points)

Development (4.1) and new location (4.2) cost seem equally important, but Earth launch cost only applies to space locations, so we give it half the weight. Therefore the weights are 14, 14, and 7 points each.

Technical Risk, Safety, and Sustainability Group (30 points)

We subjectively rate location (6.1) and population (6.2) risk more highly than the remaining factors, and thus give them 7.5 points each. Technical risk (5.1), biosphere security (7.1), and survivability (7.2) then get 5 points each.

Resulting Evaluation Criteria

From the above discussion, we can now make a table of the resulting evaluation criteria to apply to our design options. Note that in some cases scores can go outside 0 to 100% range if the parameter is outside the expected range. Like all parts of the conceptual design, this may get revised by later work.

Criterion	Weight (points)	Scoring Formula (percent)	Notes
1.2 Program Scale (per location)	3.0	ln(average population per location/100) x 25%	Population is final design size for location after growth
1.2 Program Scale (total all locations)	4.5	ln(total population all locations/5000)x25%	Population is total design size after growth
2.1 Number of locations (count)	3.75	actual count of locations > minimum size @ 1% each	Minimum size = final size/years to grow to final size
2.1 Number of locations (range)	3.75	steps in environment, time, and distance range @ 0.5% each	10 parameters and definition of steps from discussion 2.1 above
2.2 Growth (rate/yr)	5.0	(equivalent % annual GDP growth of all locations -2.5%) x 10	internal production valued as if sold at market rates
2.3 Improved Technology (local resources)	1.0	% of local resources from program locations	by kg (mass) or Joules (energy)
2.3 Improved Technology (self production)	1.0	% of finished products from program locations	by economic value
2.3 Improved Technology (cyclic flow)	1.0	% of location mass flows reused	includes propellants, but not production for growth or sale
2.3 Improved Technology (automation)	1.0	% reduction human labor hours	relative to current technology
2.3 Improved Technology (autonomy)	1.0	% required labor and control from within locations	based on necessary location functions
2.4 Quality of Life (GDP)	5.0	(equivalent GDP - $20,000)/1600	includes value of internal production and labor
2.6 Resources (surplus)	5.0	ln(material & energy output/internal use)/ln(2) x 25%	over program life cycle.Clip at -100%
4.1 Total development cost (Earth)	14.0 - S	(avg unit cost/total development cost) x 1000%	S = 14 x (space/total) development cost
4.1 Total development cost (Space)	S	(avg unit cost/total development cost) x 100%	see above for S
4.2 New Location Cost (Earth)	14.0-S2	[(ln(0.25xUS capital per person/location cost))/ln(2) x 25%]+100%	$200K includes land value for US capital. S2 see below
4.2 New Location Cost (Space)	S2	[(ln(0.5 US capital per person/location cost))/ln(2) x 25%]+100%	S2 = 14 x (people in space/total in program)
4.3 Earth Launch Cost ($/kg)	7.0	log($1600/(LEO transport per total system mass)) x 20%	total mass includes local space resources
5.1 Technical Risk Allowance (%)	5.0	(50% - technical uncertainty allowance) x 2	includes performance and design uncertainty
6.1 New Location Risk (relative)	7.5	[ln(0.25x general casualty risk/location risk)/ln(2) x 25%] +100%	casualty risk includes life and property
6.2 Population Risk (relative)	7.5	(% reduction to general population risk) x 5	from natural and program causes. Increased risk not allowed.
7.1 Biosphere Security (species-locations)	5.0	[(log(species maintained outside natural range x locations)) - 1] x 20%	in vivo or stored, humans are a species
7.2 Survivability (relative)	5.0	(% compensation for critical risks) x 5	includes all civilization level risks

Total	**100**	**Sum partial scores x weight from each line above**	

continue to page 5

Section 5.1: Human Expansion (page 5)

back to page 4

Functional Analysis

In a conceptual design such as this we want to establish one or more program concepts. The concepts describe in a general way how the program will operate. Then we begin a process of breaking down this general description into more detailed steps called **Functions**. Functions generally transform a set of inputs into a set of outputs. These inputs and outputs are called **Flows**, and may include any kind of entity: humans, data, hardware, energy. Functions do not define how the transformations happen, just that they occur. The how is embodied in a design solution or alternative, which will come later

Human civilization already exists, and is expanding in measures of population, energy use, and GDP. There are already space programs that also exist. So the first program concept, which we will call the **Existing Baseline** is to simply continue current activities without adding anything new. Any other program concepts will need to score better than the baseline to justify going forward with them. We described an initial program concept based on a particular design approach on page 2. To do a thorough job, we should examine other design approaches to see if they lead to promising concepts. Given that a single author is contributing this section at present, this will need to be limited in scope to examining approaches already proposed by others. There is an opportunity here for other individuals or teams to add to this work with additional approaches and concepts. The important thing is to compare alternatives on a fair basis, using similar goals, evaluation criteria, technology levels, and cost estimating methods.

Current Parameter Status

Some of our criteria for comparing programs are scored relative to features of existing civilization, so as a starting point we will identify the current status of these parameters. The initial version of this functional analysis is being written in late 2012, so we will use the start of 2013 as "current". If the entire conceptual design process takes a long time, the information should be updated to a new current date. 2015 is suggested for the next update.

Environment Ranges

The Temperate range is defined as where 90% of people currently live, and more difficult and extreme environments are respectively > 10% and 20% beyond that. Analysis of where people live vs environment conditions is being developed separately in the **Sec 5.2 - Environment Ranges Design Study** This is being done in two steps due to the mass of detail in mapping out where people live. The following table lists initial estimates for the boundaries of the Temperate range, and the inner boundary values for the difficult and extreme ranges. When the second part of the Environment Range study is complete, these will be updated.

Parameter	Units	Temperate Low	Temperate High	Difficult Low	Difficult High	Extreme Low	Extreme High
Environment Temperature	degrees K (C)	260 (-13)	310 (37)	234 (-39)	341 (68)	208 (-65)	372 (99)
Water Supply	(meters,tons)/m^2/year	0.25	2.5	0.225	2.75	0.20	3.0
Environment Pressure	kPa gas	80	100	72	110	64	120
Ground Pressure	MPa liquid/solid	0.25	2.0	0.225	2.75	0.20	3.0
Energy Suplly	W/m^2	150	900	135	1000	120	1100
Gravity Level	m/s^2	9.79	9.81	8.8	10.8	7.8	11.8
Radiation Dose	mSv/year	1.0	13	0.9	14.3	0.8	15.6
Ping Time	seconds	8 ms	100 ms	7.2 ms	110 ms	6.4 ms	120 ms
Travel Time	(hours, days)	8 hr	48 hr	7.2 hr	53 hr	6.4 hr	58 hr
Stay Time	years	25	70	22.5	77	20	84
Transport Energy	MJ/kg	0.22	2.25	0.2	2.5	0.18	2.75

Existing Baseline Program

Our existing baseline can be defined by looking at world development trends and existing space programs to the point that uncertainties in items like GDP or new technology are larger than 50% of our estimates. At that point the future is too uncertain to project reasonably, therefore we stop. Future updates to this kind of conceptual design study, done in 5 or 10 years time, can then project forward to a new uncertainty horizon. Other program concepts will also have a time horizon bounded by uncertainties, and should explicitly identify what those limits are. In examining world development and existing space programs, we do not need to examine their every detail, but only the parts that affect the scoring in the evaluation criteria we have chosen.

World Development Trends

Existing Space Programs

Evaluation Score

We take our previous table of evaluation criteria, and apply them to the current baseline:

Criterion	Weight (points)	Score	Notes
1.2 Program Scale (per location)	3.0	118%	Based on average population of 11,250 per 27,000 US places
1.2 Program Scale (total all locations)	4.5	163%	Based on 3.5 million living >10% beyond normal environment ranges
2.1 Number of locations (count)	3.75	311%	
2.1 Number of locations (range)	3.75	22%	
2.2 Growth (rate/yr)	5.0	5%	Based on 3% growth projection from Conference Board
2.3 Improved Technology (local resources)	1.0	20%	Based on location food and manufacturing of 20%
2.3 Improved Technology (self production)	1.0	20%	
2.3 Improved Technology (cyclic flow)	1.0	5%	
2.3 Improved Technology (automation	1.0	0%	No reduction from baseline since this is the baseline
2.3 Improved Technology (autonomy)	1.0	20%	
2.4 Quality of Life (GDP)	5.0	-6.25%	World GDP/capita in 2012 of $10,000 US
2.6 Resources (surplus)	5.0	0%	World has no net surplus by definition
4.1 Total development cost (Earth)	14.0 - S	217%	based on 2.3% OECD R&D/GDP
4.1 Total development cost (Space)	S	50%	S = 0.0007
4.2 New Location Cost (Earth)	14.0-S2	115%	US Capital = 200K/person, world = 33,000/person
4.2 New Location Cost (Space)	S2	-134%	S2 = 14 x 10^{-9}, space = $66 million/person
4.3 Earth Launch Cost ($/kg)	7.0	-9.3%	Based on Falcon 9 cost
5.1 Technical Risk Allowance (%)	5.0	100%	no uncertainty since it is existing baseline
6.1 New Location Risk (relative)	7.5	50%	relative risk is 1.0 for world as a whole
6.2 Population Risk (relative)	7.5	0%	no change, since it is baseline
7.1 Biosphere Security (species-locations)	5.0	80%	counts zoos and seed banks
7.2 Survivability (relative)	5.0	0%	no change, since it is baseline
Total	**100**	**76.7**	Estimated total score

New Program Concept

Our new program concept was first stated in a general way on page 2. The basic approach was to enable expansion by developing advanced technologies, then build a series of new locations in more difficult environments, and within each location increase in size and technical performance. The obvious functional breakdown is then first by level of technology, which defines what is possible to build, then the locations, defined by a set of environment conditions. As better technology becomes available, the locations are upgraded to targeted size and performance levels. A "location" is a general environment, such as tropical ocean or Low Earth Orbit. Selection of specific sites is left to later. To determine how many phases of upgrades are needed, we will optimize based on what can be reached in a first phase, and what size steps are reasonable. As a goal, we think a phase should provide at least a 10 point increase in evaluation score, to be significant enough to implement.

The primary functions within a location include production, habitation, and transport capacities. These allow it to support people at the location, and interact with the rest of civilization. We will develop functional flow diagrams to model these elements and the flows that connect them to each other and outside the program.

Expansion Phases

Our partial evaluation score for existing civilization is about 75 points, so we will attempt to formulate a location concept that yields a total score of 85 for Phase I. Later will will adjust and optimize this concept. For now we just want a single example to start with. Later phases will try to increase in steps of 10 to 105 points in Phase III. We don't expect to be able to plan beyond Phase III at this early date, since implementing that far in the future involves too much uncertainty.

Develop Technology (Phase 0)

Before we implement the first phase, we have to develop sufficient technology to reach the desired program goals for that phase. This includes doing conceptual and preliminary design, developing new technologies, and building prototype systems to demonstrate performance. This preliminary phase we will call Phase 0. Assuming this phase shows enough gains over the current baseline, we then proceed to the next phase.

Conceptual Design

Preliminary Design

New Technologies

It is not necessary to duplicate work for technologies which already have large efforts in progress. Thus electronics, for example, is not an area we would put much effort into. It is already a major industry with high levels of funding for technology development. Instead we will apply efforts to areas specific to the program, and which are not getting sufficient attention. The selection of which technologies to work on first will depend on ranking their relative potential impact, development difficulty, and timing of the need for it.

As a new technology reaches a sufficient level of improvement, it will migrate from this task to prototyping. If it performs well enough, it will then move to one of the later implementation phases. If more progress appears possible, a given technology goes back into the development cycle, and the amount of effort based again on ranking vs other technologies. Thus Phase 0 does not end once Phase I starts, but continues as long as there is sufficient improvement possible, periodically feeding new improvements to later locations and phases.

Prototype Scaling

It is less expensive to build smaller prototypes to try out new technologies and demonstrate performance, so we will establish a series of mass and linear scales derived from the "full scale" level of supporting 75 people/year. Not everything can be prototyped this way, but it will be used where effective. Habitation is an obvious example where scaling in all dimensions is not feasible, although scaling in area is possible. The scale steps get smaller as size goes up due to increasing cost. Consistent scaling steps helps ensure different items will work together when integrated. Initial sizes of prototypes and which scale steps to use will be determined for each technology. There is an opportunity to sell copies of the smaller scale elements, or do initial production and operations with them, as ways of generating income for further growth.

The scales are as follows:

- **1/10 Scale -** This is 10% linear, or 0.001 mass and volume, or support 0.075 people/year capacity
- **1/5 Scale -** This is 20% linear, or 0.008 mass and volume, or support 0.60 people/year capacity
- **1/3 Scale -** This is 1/3 linear, or 0.037 mass and volume, or support 2.75 people/year capacity
- **1/2 Scale -** This is 50% linear, or 0.125 mass and volume, or support about 9-10 people/year capacity

- **3/4 Scale** - This is 75% linear, or 0.422 mass and volume, or support about 30 people/year capacity
- **Full Scale** - This is 100% linear, or 1.000 mass and volume, or support 75 people/year capacity

Later growth may require larger prototypes, and multiple copies of a given size may always be used for more capacity

An example of scaling is reducing a 36 inch x 21 ft capacity, 30 hp electric commercial sawmill down to 0.5 hp electric (a factor of 60 in power), and reducing log capacity by 2.5 in each axis (a ratio of 15.625 in log volume), giving a total reduction of 937.5, roughly the 0.001 mass and volume scaling for the smallest size. In US terms the log capacity is 15" diameter by 8 ft long or 12" diameter by 12 ft long. The smaller motor is suitable for slower production rates, and the smaller size is suitable for home use. 600 lb log weights can be handled by a single person using leverage at each end singly

Develop Phase 1 Locations

The following program parameters result by working backwards from an 85% score for each evaluation criterion, or 10% higher than the baseline. They are goals, the actual values will be found at the end of the conceptual design.

- **1.2 Program scale:** Average population/location = 3000, total population = 150,000. Number of locations = 50. We will assume an inverse size distribution where size for the nth largest location is 1/n times the largest location. For 50 locations the total is 4.5 times the largest. This comes from **Zipf's Law**, an empirical observation for city sizes. This gives a smallest location of 660 and a largest of 33,000.
- **2.1 Number of Locations:** Actual count of locations is 50, giving 50% score. We compensate by increasing range of environments to 240 steps, giving 120% score.
- **2.2 Growth:** 11% per year, giving 9 year time to completion, and 75 people as minimum location design size.
- **2.3 Improved Technology:** 85% direct values for local resources, finished products, recycled fraction, automation, and autonomy.
- **2.4 Quality of Life:** Equivalent GDP (counting internal production) = $156,000 US
- **2.6 Resources:** 10.5 x internal materials and energy over life cycle, or 950% surplus.
- **4.1 Total Development Cost:** Allow 11.7x unit cost on Earth, 1.2x in space. Allows $890,000 development/person for temperate location, +10% per environment step compounded for more difficult locations, and ln(size) for increased size of same environment.
- **4.2 New Location Cost:** Per person = $76,000 Earth, Space = $152,000.
- **4.3 Earth Launch Cost:** $23/kg including space resources factor. Nominal split is $150/kg actual launch cost and 15% non-space resources factor
- **5.1 Technical Risk:** 7.5% technical uncertainty
- **6.1 New Location Risk:** Allow 38% casualty risk for new locations
- **6.2 Population Risk:** 17% reduction to population risks
- **7.1 Biosphere Security:** 178,000 species x locations maintained outside natural range
- **7.2 Survivability:** 17% compensation for critical risks

Schedule

We assume a notional schedule which ramps up gradually. Assuming a "zeroth" location for prototyping we allow 6.5 years for Phase 0 to develop the technologies, after which the first Earth location starts at a scale of 75 people and grows by 75 per year to 660. Annual population growth increases by 11% per year, so added locations are started when enough margin over first location exists, until all 50 locations are built, nominally ~50 years.

Technology development is assumed to continue for 6.5 more years to initiate Phase II, and 7 more years to initiate Phase III, so later sites will use upgraded technology, and older sites retro-fitted to improved levels. Budget for prototype work is $66 million, resulting in ~36 resident capacity

Later Phases

At this point it is too early to try and define later phases in much detail, aside from setting evaluation scores of 95 and 105%.

[Following text saved from Section 4.1:]

Level 2: Phases

The program phases are defined for now by aiming for a 10 point increase in evaluation score per phase. Early estimates of the existing baseline indicate a score of 20 points, so Phase I would aim for a score of 30, Phase II for 40 points, and Phase III for 50 points. Once the design alternatives are better understood, and what performance is feasible, the number and spacing of phases may be changed later.

Numerical Goals

Phase I

Phase I aims at a 10 point improvement over the existing baseline. The baseline is estimated to score 20 points, thus this phase aims at 30 points. The exact program parameters to reach this score depend on many lower level choices still to be made, and technology still to be proven. We know that this phase will involve some number of Earth and Near Space locations, and some level of improved technology.

Phase II

Phase II would aim at a further 10 point increase in program score. This will likely require more technology development, and because of the time span since the preliminary phase, some redesign and upgrade of Phase I elements will likely be needed.

Phase III

Phase III is currently aimed at a program score of 50 points. Because it is further out in time and difficulty, this phase is left as a more preliminary concept, more to guide the direction of the earlier phases. Unanticipated new technology is likely to affect designs this far in the future. Phases after this one (IV+) are therefore reserved for future design work.

[End saved text]

Locations List

This list is tentative pending completion of the environment ranges study

- **Temperate Earth** - Our first environment is within the middle 90% range of current civilization. The reason is to first develop improved technology such as seed factories and cyclic flows, where it is easiest to do the work, and where it will have the widest immediate application.

- **Non-Temperate Earth** - This group has one or more parameters outside the Temperate range, but not reaching the Difficult level. They are named by whichever parameter is most out of range. Many combinations are possible.

  - ***Hot Locations*** - These are where summer daytime highs exceed 310K (37C). Death Valley, California, generally considered the hottest place on Earth, reaches 47C for average summer daily high, so no place on the surface reaches the difficult level of 341K (68C) except for volcanic areas. The TauTona gold mine in South Africa, which reaches 4 km below the Earth's surface, has a rock face temperature of 60C, so it also does not reach the difficult level.

  - ***Cold Locations*** - These are where winter nighttime lows are below 260K (-13C). Vostok Station, Antarctica, is considered the coldest place on Earth, and reaches an average winter low of 201K (-72C). This is below the extreme low threshold of 208K (-65C), so the Earth includes Non-Temperate, Difficult, and Extreme Cold Locations.

  - ***High Water Locations*** - Nominally these are areas that get more than 2.5 meters (100 inches) of rainfall or other fresh water sources, such as river flow, per year. Since the maximum water supply reaches about 10 meters/year, there are about 14 steps above the Temperate range on Earth, in 10% compounded increments. Significant tropical areas fall into the Non-Temperate range from rainfall alone, and some areas near major rivers reach the highest values.

- ***Low Water Locations*** - Nominally these are areas that get less then 0.25 meters (10 inches) of rainfall or other fresh water sources/year. The driest location on Earth, the Atacama Desert in Chile gets as little as 0.001 meters/year, so there are 10 steps in dryness below the Temperate range, in 0.025 m/yr increments. Significant parts of the Sahara Desert fall into the lowest step.

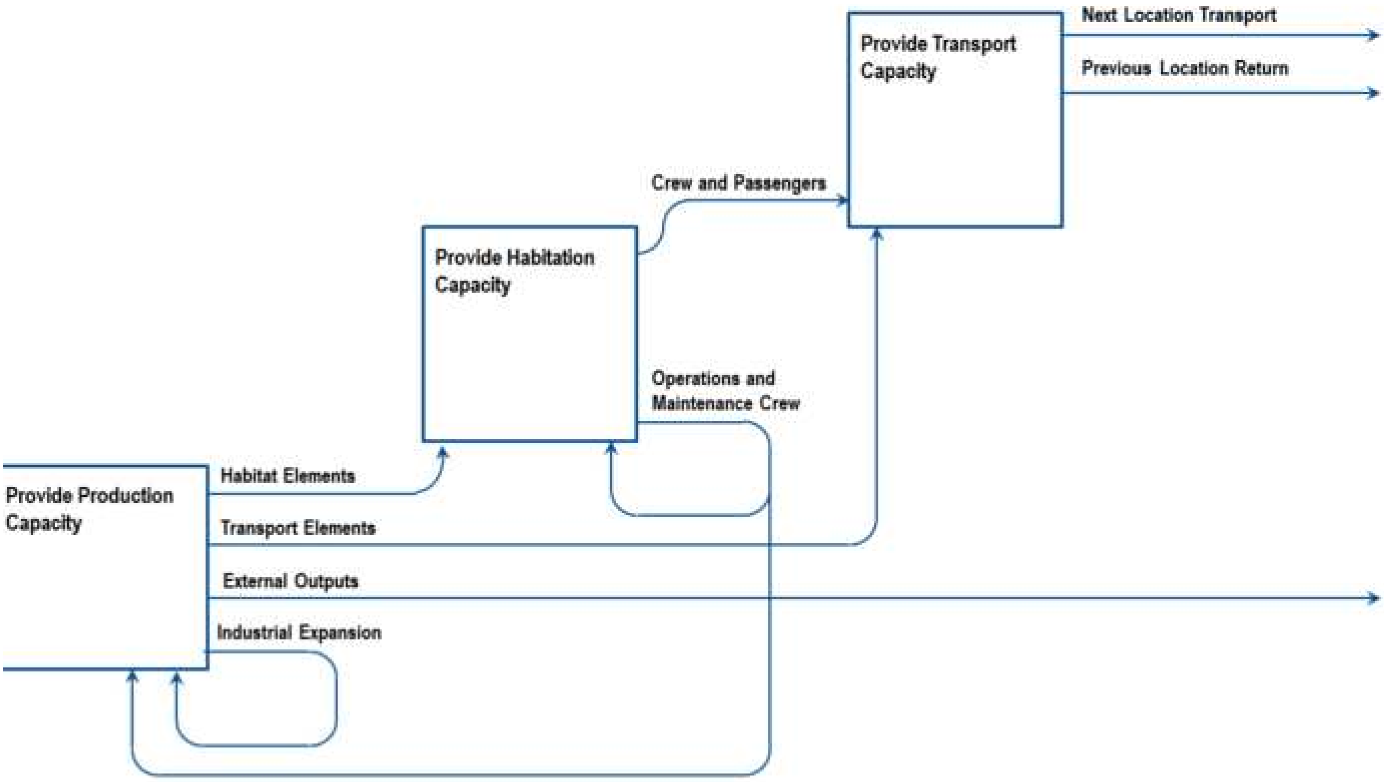
- **Difficult Earth Locations**- This group of locations push one or more environment parameters more than 10% beyond the temperate range.Based on our environment parameters, we can start to identify such locations, and then combine ones where multiple parameters can be addressed at once.All the ranges are based on what the upper and lower 5% of current population live with:

- **Extreme Earth Locations**- This group includes as many locations as needed to push environment parameters to the limits of practicality.Some parameters may have no practical use beyond a limiting value even if conditions exist beyond them.

- **Near Earth Space**- These start with the lowest useful Earth orbits at about 200 km altitude and extend upwards to 10% beyond Earth escape energy.

- **Distant Space**

Functions



eneralized development model.

Figure 5.1-1 is a very preliminary diagram showing the functional elements and flows for a generic location. Functions must trace back to at least one program goal or requirement, otherwise they are unnecessary. This can be either a direct reference to a source, or an indirect derivation by analysis.For the elements in this diagram at least one source each that justifies their inclusion are:

- **Diagram as a whole** - Multiple locations will make up the total program, so the entire diagram is an element in a higher level diagram. The need for locations comes from 1.1 Program Goal - "...a series of new locations...".
- **Provide Production Capacity** - comes from 2.3 Improved Technology - "...increase the levels of self production..."
- **Provide Habitation Capacity** - comes from 1.2 Program Scale - "...permanently supporting at least 95,000 humans total among new Earth locations and at least 2,000 humans per new space location." The physical support to live in a given location we will call Habitation.
- **Provide Transport Capacity** - comes indirectly from 1.1 Program goal "...expand human civilization..." The new locations are not cut of from existing civilization or each other, therefore they need capacity to transport people and supplies in, and products out. It comes more directly from "...a series of new locations...", since the mere existence of new locations requires transport to set them up.

Flows into and out of the diagram, and between functions can contain any kind of hardware, software, data, or people. Later analysis will define exactly what each flow contains, but they must follow the rule that flows are conserved. This comes from the physical fact that items do not appear from or disappear into nothing. Thus dividing or combining flows must sum to the same totals on both sides, and so must inputs and outputs to a function (although a function may convert the types of flows). Conservation of flows ensures that all inputs and outputs of a system are considered and accounted for. In this preliminary diagram we merely identify some of the major flows.

To continue the functional analysis, we break down the three top-level functions into lower level elements. Partitioning functions into more detailed ones creates logical boundaries inside a larger system. This then identifies flows which cross the internal boundaries, and creates simpler elements to design. The lower level functions should have an internal coherence or relatedness. Since the partitioning is logical, and not physical, it can be done in different ways, and often is to develop alternate designs. A good understanding of the nature of the system is very helpful in developing the lower level functions, which in turn may require specialist knowledge.

As a start, we can create one list of lower level functions drawn from past experience. They will likely strongly interact with each other, with many flows between the sub-functions, so a flow diagram may be too complicated to use. We will consider a table or spreadsheet instead. We also identify categories of inputs to and outputs from the location as a whole. A location is connected physically to its environment, and interacts with other locations and civilization as a whole, so flows crossing the location logical boundary are part of the analysis. The inputs and outputs will then later be divided among the more detailed functions.

Location Inputs

- Energy Sources
- Food Sources
- Water Sources
- Parts and Materials Supply
- Tools and Machines Supply
- Land Inputs
- Human Inputs
- Money Inputs
- Information Inputs

Location Outputs

- Surplus Energy
- Surplus Food
- Surplus Water
- Surplus Parts and Materials
- Surplus Tools and Machines
- Surplus Land
- Surplus Humans
- Money Outputs
- Information Outputs
- Waste Outputs

Production Functions

- Control Location
- Supply Power
- Extract Materials
- Process Materials
- Fabricate Parts
- Store Inventory
- Assemble Elements
- Grow Organics

Habitation Functions

- Protect From External Environment
- Control Internal Environment
- Provide Food
- Maintain Health
- Provide Personal Items
- Provide Information

Transport Functions

- Deliver Bulk Cargo
- Deliver Delicate Cargo
- Deliver Humans

[Parked Content from Section 4.1]

Initial Program Analysis

Program Requirements

The initial set of program requirements were developed by carefully looking at the program goals and benefits, general systems engineering experience, and natural and human constraints. We divided the general goals into more specific statements with measurable parameters. We also looked at our ideas for approaching the design, and the first list of elements to be included in the program, to see if they yield any requirements. We combined and formalized the resulting statements to create a first draft of the top level requirements. Some of the numerical values are arbitrary, but we need to set something as a starting point, which can be adjusted later as the design evolves.

1. Objectives

- **1.1 Program Goal** - The program shall expand human civilization to a series of new locations with increasingly difficult environments and distance.
- **1.2 Program Scale** - Expansion shall be demonstrated by permanently supporting at least 95,000 humans total among new Earth locations and at least 2,000 humans per new space location.
- **1.3 Choice** - Specific locations and their internal organization, function, and operation shall be chosen by program participants and location residents within the limits of design constraints.

2. Performance

- **2.1 Number of Locations** - The design shall maximize the number of new locations, where new is defined by at least a 10% increase in an environment parameter or distance measured in time or energy terms.
- **2.2 Growth** - Each location shall increase the capacity for production, habitation, and transport in a progressive manner.
- **2.3 Improved Technology** - Locations shall increase the levels of self-production, cyclic flows, and autonomy in a progressive manner.
- **2.4 Improved Quality of Life** - Completed locations shall provide an improved physical and social quality of life relative to the upper 10% of Earth civilization.
- **2.5 Data** - The program shall collect and disseminate [TBD] data about the Earth's environment, surrounding space, and objects therein.
- **2.6 Resources** - The program shall output a life cycle surplus of at least 100% of internal material and energy resource needs.

3. Schedule

- **3.1 Completion Time** - The expansion to a new location shall be completed before expected progress in technology indicates a re-design is required.

4. Cost

- **4.1 Total Development Cost** - The total program development cost for new technology and hardware designs shall be less than 50 times the unit cost on Earth, and 5 times the unit cost in space of the hardware.
- **4.2 New Location Cost** - The peak net project cost for a new location shall be less than 50% of the expected long term net output.
- **4.3 Earth Launch Cost** - The program shall progressively lower the Earth launch cost component of total system cost, with a goal of $0.08/kg of total system mass.

5. Technical Risk

- **5.1 Risk Allowances** - Program designs shall include allowances for uncertainties and unknowns in knowledge, performance, failure rates, and other technical parameters. New designs with higher risk can be included in program plans, but a process shall be included to resolve the risk, and an alternate design with lower risk maintained until resolved.

6. Safety

- **6.1 New Location Risk** - New locations shall progressively lower internal risks to life and property, with a goal of significantly lower risk than the general population.
- **6.2 Population Risk** - The program shall significantly reduce natural and human-made risks to the general population, including external risks created by the program.

7. Sustainability

- **7.1 Biosphere Security** - The program shall increase biosphere security by establishing alternate biospheres and long term storage of biological materials.
- **7.2 Survivability** - The program shall design for the long term survival of life and humanity from changes to the Earth which will render it uninhabitable and depletion of critical resources.

8. Openness

- **8.1 Open Design** - Technology and design methods developed within the program shall be open for others to use. Specific instances of a design and produced items may be proprietary
- **8.2 Access** - Development of a new location shall not prevent reasonable access for transit or to unused resources.

Evaluation Criteria

Setting discrete program requirements like the ones listed above are unlikely to be the optimum values, and do not help in choosing among design alternatives. For those purposes we choose parameters to measure our evolving design and guide it to the preferred result. We identify these parameters by again carefully looking at all the work done so far, and selecting the ones most important at the program level. After selection, we then scale and adjust their relative importance to each other so that a score can be determined for each design option or variation. Our resulting criteria and how they are scored is as follows:

Criterion	Weight (points)	Scoring Formula (percent)	Notes
1.2 Program Scale (per location)	3.0	ln(average population per location/100) x 25%	Population is final design size for location after growth
1.2 Program Scale (total all locations)	4.5	ln(total population all locations/5000)x25%	Population is total design size after growth
2.1 Number of locations (count)	3.75	actual count of locations > minimum size @ 1% each	Minimum size = final size/years to grow to final size
2.1 Number of locations (range)	3.75	steps in environment, time, and distance range @ 0.5% each	10 parameters and definition of steps from discussion 2.1 above
2.2 Growth (rate/yr)	5.0	(equivalent % annual GDP growth of all locations -2.5%) x 10	internal production valued as if sold at market rates
2.3 Improved Technology (local resources)	1.0	% of local resources from program locations	by kg (mass) or Joules (energy)
2.3 Improved Technology (self production)	1.0	% of finished products from program locations	by economic value
2.3 Improved Technology (cyclic flow)	1.0	% of location mass flows reused	includes propellants, but not production for growth or sale
2.3 Improved Technology (automation	1.0	% reduction human labor hours	relative to current technology
2.3 Improved Technology (autonomy)	1.0	% required labor and control from within locations	based on necessary location functions
2.4 Quality of Life (GDP)	5.0	(equivalent GDP - $20,000)/1600	includes value of internal production and labor
2.6 Resources (surplus)	5.0	ln(material & energy output/internal use)/ln(2) x 25%	over program life cycle.Clip at -100%
4.1 Total development cost (Earth)	14.0 - S	(avg unit cost/total development cost) x 1000	S = 14 x (space/total) development cost
4.1 Total development cost (Space)	S	(avg unit cost/total development cost) x 100	see above for S
4.2 New Location Cost (Earth)	14.0-S2	[(ln(0.25xUS capital per person/location cost))/ln(2) x 25%]+100%	includes land value for US capital. S2 see below
4.2 New Location Cost (Space)	S2	[(ln(0.5 US capital per person/location cost))/ln(2) x 25%]+100%	S2 = 14 x (people in space/total in program
4.3 Earth Launch Cost ($/kg)	7.0	log($1600/(LEO transport per total system mass)) x 20%	total mass includes local space resources
5.1 Technical Risk Allowance (%)	5.0	(50% - technical uncertainty allowance) x 2	includes performance and design uncertainty
6.1 New Location Risk (relative)	7.5	[ln(0.25x general casualty risk/location risk)/ln(2) x 25%] +100%	casualty risk includes life and property
6.2 Population Risk (relative)	7.5	(% reduction to general population risk) x 5	from natural and program causes. Increased risk not allowed.
7.1 Biosphere Security (species-locations)	5.0	[(log(species maintained outside natural range x locations)) - 1] x 20%	in vivo or stored, humans are a species
7.2 Survivability (relative)	5.0	(% compensation for critical risks) x 5	includes all civilization level risks

Total	**100**	**Sum partial scores x weight from each line above**	

Retrieved from "https://en.wikibooks.org/w/index.php?title=Space_Transport_and_Engineering_Methods/Combined_Systems5&oldid=3172481"

Section 5.2: Environment Ranges

This study is preliminary as of Nov 2012

Introduction

This design study is part of the Human Expansion (HE) program conceptual design. One of the HE program goals is to expand civilization into more difficult locations, while at the same time improving levels of technology and quality of life. The goal in this study is to establish location parameter ranges for where people live now, and then from that derive what difficult and extreme ranges would be for future locations. We set the middle 90% of where people live now as the existing range for each parameter, with 5% living at each extreme. The middle 90% is described as **Temperate**, and levels more than 10 and 20% beyond the Temperate range we call **Difficult** and **Extreme** respectively.

The study will be performed in two parts. The first part will make make initial estimates so that other parts of the HE program design can start to use the information. The second part will survey where people live in more detail and generate improved values. The HE program as a whole is using the start of 2013 as a baseline to reference against. The **US Census Bureau** projects population at the start of 2013 to be 7,062 million, thus 5% is 353 million. The more detailed population analysis will draw from world statistical data.

Initial Estimates

Seven environment parameters and four time and distance parameters were chosen in the HE Requirements Analysis to determine what makes a difficult or extreme location. In the list of the parameters that follows, we first define what the parameter is, then discuss how we make our estimate for the 90% Temperate Range. We also discuss what the practical limits on Earth are for the Extreme ranges. Once the practical limits on Earth are reached, or they exceed what is found in space, more difficult locations in space are used.

Environment Temperature

Definition - On Earth, this is the winter daily air temperature lows, and summer daily highs in Kelvin and (Celsius). For space locations this is the equilibrium temperature of a 50% gray body with one side facing the Sun and the opposite side facing away

Estimate - On inspection of a world map, many more people live in far north locations than far south. Northern Europe, Russia, parts of Tibet, Northern China, and Canada may supply 350 million cold climate residents. Many people live in equatorial/hot climates, but the extremes are seen in dry climates where water vapor does not moderate daily temperature ranges. We make a rough estimate of 260 K (-13 C) as average January daily lows and 310 K (37 C) as average July daily highs as the limits of the Temperate range.

Limits - Ten percent, or 25-30K, beyond the temperate limits is a significant range for human climate, so Earth locations are likely to to have only one or two very cold and no hot difficult surface climates. Deep underground might reach extreme hot levels in certain locations.

Water Supply

Definition - This is fresh water supply from rain and snow fall, flows from rivers and ice, and air moisture in meters depth/location area/year. Since the density of fresh water is 1 ton/m^3, this is equivalent to tons/m^2. Salt water does not count as fresh water supply, although evaporation does. Underground aquifers that represents flows from rain or water transport count, but not amounts merely drawn from storage, since that is not sustainable. Many people live near rivers, which supply high fresh water flows. We count available river flow distributed over the local land area, so a high range location above that may be difficult to find. Dry locations with low water flow are common.

Estimate - We use values of 0.25 and 2.5 meters on Earth as the Temperate range based on rainfall definitions of desert and wet climates. This needs more support from climate data.

Limits - On Earth, the lowest rainfall location, the Atacama Desert receives an average of 0.001 meters of rainfall/year. We can get an upper estimate by assuming the Amazon River flow of 200,000 cubic meters/second is allocated to the nearest 10% of the drainage basin to the main river and tributaries. Given a total area of 7 million km^2, that gives a water supply of 9 meters per year. The highest rainfall locations give about 11 meters/year. So we will adopt 10 meters/year as the extreme upper value. Space locations would range from essentially zero in the inner Solar System, to moderate when significant numbers of ice or hydrated bodies pass within reach, to very high for ice-covered satellites or bodies.

Atmosphere Pressure

Definition - The average location gas pressure in kPa.

Estimate - On Earth nearly everyone lives in a fairly narrow range of pressures. We will assume for now the nominal range is from 80 to 100 kPa (sea level to 2000 m altitude).

Limits - On Earth the lowest surface pressure is presumed to be at the highest point, Mt. Everest, at +8,850 m elevation. The highest pressure is presumed to be at the bottom of the LaRonde Gold mine in Quebec, which reaches 3 km below a surface elevation of 300 m, for a net height of -2,700 m. The pressure range is thus from 31.5 to 138 kPa. To go beyond these limits would require artificial structures, or ambient pressure underwater locations. In space, local pressure can range from complete vacuum to many times Earth values on Venus or the Gas Giants.

Ground Pressure

Definition - On bodies with gravity this is the foundation design load in MPa or exterior water or rock pressure for below surface locations. In both cases this amounts to a boundary pressure on location structures. Low range would be surface water structures or swamps, and high range would be tall buildings or below surface locations.

Estimate - We make a rough estimate of the Temperate range by referring to allowable soil bearing pressures from soft clay up to hard bedrock. Dropping extreme values, these give a range from to 0.25 to 2.0 MPa.

Limits - For low range limit, we can assume a shallow boat or raft foundation with a depth of 1 meter or less. This gives a load of 0.01 MPa. For high range limits, assuming a floor + dead load of 1000 kg/m^2 per floor, a 100 story skyscraper might impose a bearing pressure of 1 MPa on the foundation base. Practical limits on tall construction have more to do with economics than with structural materials, so taller buildings are possible.

Since the depths of the ocean and deep underground can reach very extreme pressures and temperatures, we set a practical limit based on the energy to reach Earth orbit (31.3 MJ/kg) compared to the column of mass which must be displaced to access such deep locations. For a 1 kg volume of ocean, which has a mass of ~1 kg, displacing a column 800 m high requires raising 8000 kg x 400 m avg height x 9.8 m/s^2 = 31.3 MJ. For continental crust, which has a density of 2.7kg/liter, a 1 kg volume then is 7.2 x 7.2 cm. Each meter of rock above this area then is then 13.925 kgA column 675 m deep then masses 9400 kg, and raising it an average of 337.5 m also yields 31.3 MJ/kg.Therefore we will consider the practical limits to be ocean depths to 800 meters and continental depths to 675 m, after which we start including space locations. Other parameters may drive us to space before depth does. The corresponding water and rock pressures are 8 and 18 MPa respectively.

Energy Supply

Definition - Flux from natural sources in W/m^2. On Earth this will mostly be from solar and wind, with water flow and geothermal in some locations.

Estimate - On Earth, average solar flux ranges from 100 to 300 W/m^2 after accounting for night, clouds, and sun angle. Wind power in the US ranges from below 100 to over 1000 W/m^2. Hydroelectric, tidal and sea currents, and geothermal are localized, so we will ignore them for this first estimate. Combining wind and solar we get a rough estimate of 150 to 900 W/m^2 as the temperate range. In space, solar flux is 1366 W/m^2 at the Earth's distance, times the percentage time in sunlight.

Limits - Any significant water or ground depth will have near zero energy supply. Peak Earth values might be at high altitudes where a combination of high winds and increased sunlight would supply high power levels. In space, peak energy level will vary dramatically with distance from the Sun.

Gravity Level

Definition - Local gravity level in meters/s^2. Space locations near large objects still has a gravity level, even though orbits may create free-fall conditions with low relative forces between system elements.

Estimate - The Earth's surface, where almost everyone lives, is entirely within 10% of the 9.80665 m/s^2 standard value.

Limits - Values more than 10% outside the Temperate range require extremely tall structures (> 320 km tall), or orbital locations. Thus this parameter is effectively fixed for Earth locations. Solar system locations can range from zero to about 200 m/s^2 very close to the Sun.

Radiation Dose

Definition - Human radiation exposure in an unprotected state from background radiation, in milliSievert/year. It should be noted that in many locations humans cannot survive in an unprotected state, but for consistency the exposure level is measured that way. In addition to background radiation, humans also get significant exposure from medical and other human-made sources, but that is not location-specific, so is not counted as part of this location parameter

Estimate - Most of the Earth's population lives in a relatively low radiation environment, but higher altitudes or high natural radioactivity areas exist.Data derived from a UN report indicate the typical range is from 1-13 mSv/yr

Limits - The city of Ramsar in northern Iran has background levels up to 135 mSv/yr due to underground concentrations of Uranium and mobile decay products. Low Earth Orbit locations range from 80-160 mSv/yr, and locations above this can reach much higher levels from trapped particle belts and Solar particle events.

Ping Time

Definition - Minimum round trip communication delay to next nearest 5% of human population, in seconds. Depending on location, this may be by wired or wireless methods.

Estimate - In theory all of the Earth is within 0.13 second ping time from any other point, but actual communications systems in place increase this value. We assume at least voice/cellphone bandwidth for communications. There are now more fixed and mobile telephone connections than there are people, so again, in theory, everyone has access to one. From Slovakia, around 5% of the world population is within about 600 km. Theoretical ping time is then 8 milliseconds (ms). The distance from Tahiti to major population centers is roughly 6600 km. Therefore best case ping time is 44 ms. Given actual routing, and fiber-optic speeds, this is likely to be more like 100 ms.

Limits - The worst case on Earth is no worse than locations like Tahiti, which require submarine cable, or satellite relay through the Iridium network. Thus the limits for Earth are the same as for the Temperate range. Space locations can range to 45,000 seconds from Earth to the Kuiper Belt.

Travel Time

Definition - Maximum one way normal travel time for humans, in hours or days, from the nearest 5% of human population.

Estimate - In dense areas such as Europe the maximum direct distance to reach 5% of the world's population is about 600 km. Allowing for actual highway routing and travel speed, we can estimate 8 hours to reach any point in that radius. Ninety percent of the world's land area can be reached within 48 hours according to the **EU Joint Research Centre** map of accessibility.

Limits - The same accessibility map places parts of the Tibetan plateau at 20 days travel to a city of any size. **Point Nemo**, the farthest point in the world from land, is about 4000 km by sea from the nearest airports. Therefore it is about 10 days travel by chartered ship. Current travel time to space is measured in months to years, mostly because trips carrying humans are rare. The actual transit time from the launch site is 2-3 days

Stay Time

Definition - Average per person stay time per location, in years. People are assumed to stay in the same location if they live and sleep in the same place on a regular basis.

Estimate - In the US, in 2010-2011, 3.9% of the population moved to a different county in one year. We will take a county to represent a "location". Therefore the average stay time is 25 years. We assume there are significant parts of the world where the average person does not move far from their place of birth, thus the average stay time in one location is their life expectancy of about 70 years.

Limits - The fastest growing counties in the US in 2011 had a growth rate of 10%, and thus an average stay time of 7 years if both growth and normal mobility are counted. Stay times in space are currently much shorter than this because there are no permanent residents, and time is limited by zero gravity and radiation exposure. We invert the time scale for space, and set a minimum stay time of 10% of Earth, or 0.7 years, as the initial range. At the other extreme, stay times are bounded by life expectancy, so are no higher than 70 years.

Transport Energy

Definition - Maximum total energy to reach a location from nearest 5% of population, by most efficient method, in MJ/kg. Includes kinetic, potential, and friction energy cost.

Estimate - For Earth, since most of the surface is near the same potential, this is mostly frictional losses for rail or ocean shipping. Rail transport consumes 330 BTU/ton-mile, or 216 J/kg-km. Assuming a total transport distance of 1000 km for dense areas, including partial road travel from nearest rail point, we get about 0.215 MJ/kg. Ships consume 375 J/kg-km. Eastern Australia is about 6000 km by ship from the nearest 5% of the worlds population, giving 2.25 MJ/kg.

Limits - Assuming a mule pulling a cart equal to its body mass for 20 days to reach inaccessible areas in Tibet, we get 1.8 MJ/kg transport energy. If we then add another 2000 km by rail to reach 5% of the world's population, this adds another 0.43 MJ/kg, for a total of 2.23 MJ/kg. Assuming transport distance of 7000 km from population centers to Tahiti, this gives 2.6 MJ/kg, which is the higher value. Low Earth orbit requires 31 MJ/kg, so there is a large step function from Earth to space locations.

Detailed Estimates

Atmospheric Temperature

We selected the seasonal daily highs and lows for this parameter as the "normal extremes" that people are adapted to, even though particular days can exceed this range. We start from the hot and cold ends of where people live, and count until we reach 5% in each. We count by countries when they are small and consist of one climate region, or by sections when they are large. If the population is concentrated in one part, we use climate data for that area.

Coldest Populate Areas

The following table of cold locations starts at the poles and highest elevations and works down until we reach 353 million population

Country, Region	Population (million)	Average Winter Lows C (K)	Notes (subtotals)
Russia, Sakha and Chukotka	1.0	-41 (232)	at Yakutsk
Canada, 3 North Territories	0.1	-37 (236)	at Pond Inlet, Nunavut
Russia, Tuva	0.3	-35 (238)	at Kyzyl
Russia, Zabaykalsky Krai	1.1	-32 (241)	at Chita
Russia, Yamalo-Nenets	0.5	-30 (243)	at Salekhard
Russia, Buryatia	1.0	-28 (245)	at Ulan-Ude
Russia, Amur Oblast	0.9	-27 (246)	at Blagoveshchensk (4.9m)
Russia, Khabarovsk Krai	1.4	-24 (249)	at Khabarovsk city
Russia, Kemerovo Oblast	2.9	-23 (250)	at Kemerovo city
Russia, Khakassia	0.5	-23 (250)	at Abakan
Russia, Khanty-Mansi	1.4	-23 (250)	at Surgut (6.2m)
Russia, Irkutsk Oblast	2.5	-22 (251)	at Irkutsk city
Russia, Tyumen Oblast	3.4	-22 (251)	at Tyumen city
Russia, Komi	1.0	-21 (252)	at Syktyvkar
Russia, Omsk Oblast	2.0	-21 (252)	at Omsk city
Russia, Tomsk Oblast	1.0	-21 (252)	at Tomsk city (9.9m)
Russia, Altai Krai	2.4	-20 (253)	at Barnaul
Russia, Novosibirsk Oblast	2.7	-20 (253)	at Novosibirsk city
Russia, Altai Republic	0.2	-19 (254)	at Gorno-Altaysk (5.3m)
Russia, Krasnoyarsk Krai	2.8	-19 (254)	at Krasnoyarsk city
Russia, Magadan Oblast	0.1	-19 (254)	at Magadan town
Greenland	0.1	-18 (255)	at Sisimiut
Russia, Arkhangelsk Oblast	1.3	-16 (257)	at Arkhengelsk city
Russia, Perm Krai	2.6	-16 (257)	at Perm city
Russia, Sverdlovsk Oblast	4.3	-16 (257)	at Yekaterinburg (11.2m)
Russia, Kirov Oblast	1.3	-15 (258)	at Kirov city
Russia, Murmansk Oblast	0.9	-14 (259)	at Murmansk city
Russia, Vologda Oblast	1.2	-14 (259(	at Vologda city
Russia, Karelia	0.7	-13 (260)	at Petrozavodsk (4.1m)
Russia, Kamchatka Krai	0.3	-10 (263)	at Petropavlovsk-Kamchatsky
Russia, Pskov Oblast	0.7	-9 (264)	at Pskov city
Russia, Leningrad Oblast	1.7	-8 (265)	at St. Petersburg
Russia, St. Petersburg	4.9	-8 (265)	City is Federal District
Russia, Kaliningrad Oblast	0.9	-4 (269)	at Kaliningrad
Argentina, Tierra Del Fuego province	0.1	-2 (271)	southernmost province

Chile, Punta Arenas	0.1	-1 (272)	Southernmost world city of any size (8.7m)
Totals	**50.3**	**-1 (272)**	Sum of above populations, and highest winter lows

Coldest Locations

These are locations where few or no people live, with extreme environment conditions:

Country, Region	Population (million)	Average Winter Lows C (K)	Notes
Antarctica, McMurdo Station	0.0	-32 (241)	Largest populated location on the continent
Antarctica, Vostok Station	0.0	-72 (201)	Coldest place on Earth

---

Section 5.4 - Open Source Space Program

This is a conceptual design stage study of the need for, justification, and design of an **Open-Source Space Program**. The Systems Engineering process can be applied to any complex system, including the design of organizations which in turn design space projects. The study both serves as a tutorial example for such design studies in general, and a useful work product. Even a negative result, that open source space programs are not needed or justified, is useful in eliminating a poor option if that turns out to be the result.

Status: the study is approx. 3% complete as of 03 August 2012.

Background of Study

Why even consider open source? Historically space projects have been carried out by the government and commercial sectors, and data have been kept private either for national security or business competition reasons. However this results in duplication of effort, and is therefore inherently inefficient. Scientific fields, where data is published openly, and open-source software, are good examples of non-duplication of effort. The question then becomes can those methods be applied to space projects? The level of technology and size of economies restricted early space projects to the largest governments at first. As technologies have improved and the general size of economies grown, the relative scale required has gone down. Open projects tend to be relatively small, so a second question is whether space projects have reached a scale that open source projects can execute them?

Past Work On This Topic

Requirements Analysis

Defining the "Customer"

The customer in the Systems Engineering process is the entity which has an unmet need or desire which a new system might satisfy. In this case the customer is assumed to be the set of humans who want to make use of the resources of space, but which cannot because it is too difficult or expensive at the moment. There are existing government and commercial space projects, but the government ones at least have shown a remarkable lack of progress in making access to space easier and less expensive. At this point the individuals who make up the customer set are mostly unidentified, so we cannot poll them to find out exactly what requirements they might have. Instead we must start with proxies to represent what they might want, and later incorporate their real requirements if we can identify them to ask.

Establish Project Requirements

Since the unmet need is easier and less expensive access to space to make use of its resources, we must ask easier than what? We will set the baseline as the current path of already existing space programs. If an open-source approach can improve on that path, it would be worth pursuing.

Establishing Project Measures

Normally for a complex task, there will be multiple parameters that measure "goodness" in different ways. In order to compare and select among alternatives these various measures are converted to a common scale.

References and Sources

A system designer should know the current state of knowledge in topics relevant to their work. There are several reasons. One is to not repeat work already done by someone else. Another is to stimulate new ideas and improvements. In addition to the references listed below and elsewhere in the book, it is very useful to know how to find additional information. Categories of information include:

- Current News (Magazines, Newspapers, Blogs)
- Archival Publications (Journals, Preprint Archives)
- Books
- Online Data (Web and other protocols)
- Technical Reports
- Product Data
- Discussion Forums

Once information is located you should record where and how you found it, to save having to find it again. There are a number of ways to do that, depending on type of media: building a personal library in paper or electronic form, bookmarking in web browsers, etc. As long as you have the ability to find the data again, the particular methods can be left to personal preference.

Current News

As of 2012, obtaining current news via paper-based sources is nearly obsolete. Internet-based electronic delivery is faster and less expensive. There are innumerable online websites, magazines, and blogs covering every topic relevant to space systems. One distinction between sites is original reporting vs re-posting of news from elsewhere. You should also consider the quality and experience of the writers and any bias they might have. We cannot list all possible news sources, and any such list would rapidly become obsolete. Therefore we will list a few good examples and encourage looking for good sources in your fields of interest.

General News

- **Google News** - General news aggregator. This means they do not write original stories, but rather link to stories from multiple other sources. Searchable and customizable, it is a good starting point to find other regular sources. Story selection for the front page is automated, but the full database indexes several thousand news sites.
- **Reddit Technology, Science, and Space Topics** - Reader submitted daily links divided by topic, with active discussions. There are many topic headings, but the three listed are the largest related to space. It is another news aggregator, but human-submitted and filtered by votes of the readers.

General Space

- **Space.com** - General news website for anything related to space.

General Technology

- **Next Big Future** - Blog for all kinds of new and future science and technology. Often covers space topics.

General Science

- **EurekAlert** - Current science and technology press releases. Often a day earlier and unmodified by journalists.

Specific Industries

- **Aviation Week** - Aerospace industry news, including space projects.
- **Automation World** - Manufacturing automation.
- **EE TImes** - Electronics engineering. This link is to the news summary.

Multiple Documents

The following sites index, search, or link to multiple documents:

Library Indexes

Two steps in finding the current state of knowledge are (1) finding what works exist on a given topic, and (2) locating a copy of the work to look at. Library indexes help with both steps.

- **Library of Congress Online Catalog** - This is an index to the largest single library in the world, including books and periodicals. Basic search is by title, author, and subject. Items are shelved and indexed by subject, using a cataloging **Call Number** system. If you use the "Call Number Browse" type of search, you can look through all the entries under that subject, which is useful for doing general research on a subject. The **LC Classification Outline** has a summary of the call number system. For space projects, the headings Q (Science) and T (Technology) are the most relevant, although other sections can be useful depending on the topic.

- **WorldCat** - This is an online database of the collections of over 10,000 libraries worldwide. If a work is not available nearby or online, a process of **Inter-Library Loan** can be used from a local library to borrow it from another library.

Online Search

In addition to well-known search engines such as **Google** and **Bing**, there are more specialized searches such as:

- **Google Scholar** - for searching scholarly works
- **Google Books** - for searching the text of books in general
- **Open Directory Specialized Search Engine List** - Links to many other search websites.

Online Repositories

Repositories contain multiple online documents which usually can be downloaded. In some cases they only index the document online and you then have to request or find a physical copy.

- **Rocket Science Library** - A collection of open source, public domain, and online works compiled in parallel with the development of this book. A list of additional works is included for ones where copyright restrictions prevent including the document itself.

- **NASA Technical Reports Server** - Access to approximately 1 million documents in NASA databases.
- **Defense Technical Information Center** - Public access to the unrestricted subset of approximately 2 million US DOD and other technical documents.
- **US Patent and Trademark Office** - Search and download patent data. Patents often contain very useful technical data.
- **arXiv.org** - Open access to over 750,000 electronic papers in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics. Operated by Cornell University Library.
- **NATO Research and Technology** - FTP server of NATO technical and conference reports. Some have relevance to space projects. It is recommended to find titles and document numbers by other search methods, since the ones here are stored only by their document numbers.
- **National Space Society Library** - Online library with multiple space references.
- **Defense Acquisition University ACC Practice Center** - Multiple documents about how the US government manages projects, including engineering methods in general, and space projects as a topic.

Reading Lists

- **Encyclopedia Astronautica Recommended Books**

Book Length Sources

Other Wikibooks

- Astrodynamics
- High School Earth Science/Early Space Exploration and High School Earth Science/Recent Space Exploration describes how far we've come --
- Colonizing Outer Space and Conplanet speculates on how far we may go, and what living on other planets may be like after we get there.

General Space Reference

- Jet Propulsion Laboratory **Basics of Space Flight**, 2012.
- USAF Air Command and Staff College, **Space Primer, 2nd Ed.**, 2009.

General Space Systems Design

- Griffen, Michael D. and French, James R., **Space Vehicle Design**, 2nd ed., AIAA Education Series, 2004.
- Hammond, Walter E., **Design Methodologies for Space Transportation Systems**, AIAA Education Series, 2001.
- Ley, Wittmann, and Hallmann, eds. **Handbook of Space Technology**, Wiley, 2009.

Propulsion

- Glenn Research Center, **Beginner's Guide to Rockets** - website designed for young students and teachers.
- Hunter, Maxwell, **Thrust into Space**, Holt, Rhinehard, and Winston, 1966. - An introductory textbook on space propulsion.
- Sutton, George P. and Biblarz, Oscar, **Rocket Propulsion Elements**, 8th ed., Wiley, 2010, ISBN-13: 978-0470080245. - This has been considered the standard reference for classical rocket propulsion. It is mostly concerned with solid, liquid, and hybrid chemical rockets, and has one chapter on electric propulsion.
- Bolonkin, Alexander "Review of new ideas, innovation of non-rocket propulsion systems for Space Launch and Flight" 3 parts, 2010, The Internet Archive:

  Part 1, Part 2, Part 3

- Mallove, Eugene F. and Matloff, Gregory L., **The Starflight Handbook** A Pioneer's Guide To Interstellar Travel, 1989. - A semi-technical introduction to interstellar flight.
- Morgenthaler, G. W.; Tobiska, W. K. "Aerospace Century XXI: Space Flight Technologies", Proceedings of the 33rd Annual AAS International Conference, Boulder, Colorado, 26-29 Oct. 1986. Published as Advances in the Astronautical Sciences, vol 64, pt 2, 1987.

Early Works

Although technology has changed drastically since these were written, physics has not, and they serve as a guide to what technical problems need to be solved in any design:

- Bureau of Naval Personnel, **Principles of Guided Missiles and Nuclear Weapons**, 1959.
- Koelle, Heinz Hermann, ed., **Handbook of Astronautical Engineering**, New York, McGraw-Hill, 1961. - This is an excellent comprehensive reference handbook representing the state of the art as of 1961. Has a chapter covering

nuclear, electric, and solar-thermal propulsion.

- Merrill, Grayson, ed. **Principles of Guided Missile Design** series, Van Nostrand, 1958. - This is a series of at least 9 volumes with different authors. Although written in the Cold War nuclear missile era, the technology is relevant to space projects. Likely much of it will need updating due to being written 50 years ago:
  - Guidance - Locke, Arthur et. al.
  - Aerodynamics, Propulsion, Structures, and Design Practice - Bonney, E.A, Zucrow, M.J., and Besserer, C.W.
  - Operations Research, Armament, Launching - Merrill, G., Goldberg, H., and Helmholz, R.H.
  - Missile Engineering Handbook - Besserer, C.W.
  - Space Flight - Ehricke, Kraft A.
  - Systems Engineering - Jerger, J.J.
  - Range Testing - Freitag, R.F.
  - Airborne Radar - Povejsil, D.J., Raven, R.S., and Waterman, P.J.
  - Automatic Flight Control - Povejsil, D.J., Kelly, A.J., Mathews, C.W., and McCourt, A.W.

Resource Use

- Faughnan, Barbara (ed.); Maryniak, Gregg (ed.) "Space Manufacturing 5: Engineering with Lunar and Asteroidal Materials", proceedings of the 7th Princeton/AIAA/SSI Conference, Princeton, New Jersey, 8-11 May 1985.

Websites

This category includes online sites where the data is in the site itself, rather than the site being a link to other documents.

General Space Websites

- **Permanent** - Organization dedicated to space development. Their website has extensive discussion on relevant topics.
- **Rocket and Space Technology** - Personal website with extensive coverage of the basics of space flight, hardware, and missions.
- **Yarchive** - An archive of space related posts from the USENET discussion system, covering many topics.
- **Centauri Dreams** - Long-running blog about deep space exploration.

Real Time Data

- **Real Time Satellite Tracking - Provides tracking simulation of Earth satellite orbits based on official tracking data.**
- **Eyes on the Solar System - 3D Simulation of Solar System spacecraft and planetary bodies beyond Earth orbit, including current, and past and future positions.**

Life Support

- **USDA Nutrient Data Laboratory** - Includes a food and nutrition database, from which food requirements for space projects can be determined.

Articles and Technical Reports

Multiple Propulsion Concepts

- Diesposti, R. S.; Pelouch, J. J. "Performance and Economic Comparison of Externally Energized vs Chemically Energized Space Propulsion", AIAA paper number 81-0703 presented at 15th International Electric Propulsion Conference, Las Vegas, Nevada, 21-23 June 1981.
- Andrews, Dr. Dana G, ed. Advanced Propulsion Systems Concepts for Orbital Transfer: Final Report, Boeing document D180- 26680 produced under NASA contract NAS8-33935.
- Forward, Robert "Alternate Propulsion Energy Sources", AFPRL TR-83-067. (NTIS AD-B088 771/1) Dec 1983, 138p. Keywords: Propulsion energy metastable helium, free-radical hydrogen, solar-pumped plasmas, antiproton annihiliation, ionospheric lasers, solar sails, perforated sails, microwave sails, quantum fluctuations, antimatter rockets.
- Wang, S.-Y.; Staiger P. J. "Primary Propulsion of Electro-Thermal, Ion and Chemical Systems for Space Based Radar Orbit Transfer", AIAA/SAE/ASME/ASEE 21st Joint Propulsion Conference, AIAA paper number 85-1477, 1985.
- Forward, R. L. "Advanced Space Propulsion Study - Antiproton and Beamed Power Propulsion", Final Report, 1 May 1986 - 30 Jun 1987, Hughes Research Laboratories, report AFAL-TR-87-070, 1987. Defense Technical Information Center #AD-A189 218. National Technical Information Service # AD-A189 218/1 PC A10/MF A01 [Quote: , goes into detail on beamed power systems including " 1) pellet, microwave, and laser beamed power systems for intersteller transport; 2) a design for a near-relativistic laser-pushed lightsail using near-term laser technology; 3) a survey of laser thermal propulsion, tether transportation systems, antiproton annihilation propulsion, exotic applications of solar sails, and laser-pushed interstellar lightsails; 4) the status of antiproton annihilation propulsion as of 1986; and 5) the prospects for obtaining antimatter ions heavier than antiprotons." Again, there is an extensive bibliography.]
- Forward, R. L. "Exotic Propulsion in the 21st Century", in Aerospace Century XXI (see Morgenthaler and Tobiska, under Propulsion above).
- Forward, Dr. Robert L., Future Magic., Avon, 1988. ISBN 0-380-89814-4. "Nontechnical discussion of tethers, antimatter, gravity control, and even further-out topics."

More References to be sorted

These should eventually end up under the appropriate section of the book:

- Kunz, K. E. "Orbit Transfer Propulsion and Large Space Systems", J. Spacecraft and Rockets vol 17 no 6 pp 495-500, Nov.-Dec. 1980.
- Poeschel, R. L. "Comparison of Electric Propulsion Technologies", AIAA paper number 82-1243 presented at AIAA/SAE/ASME 18th Joint Propulsion Conference, Cleveland, Ohio, 21- 23 June 1982.
- Jones, R. M.; Kaplan, D. I.; Nock, K. T. "Electric Propulsion Systems for Space Stations" AIAA/SAE/ASME 19th Joint Propulsion Conference, AIAA paper number 83-1208, 1983.
- Jones, R. M. "Space Supertankers: Electric Propulsion Systems for the Transportation of Extraterrestrial Resources" AIAA/SAE/ASME 20th Joint Propulsion Conference, AIAA paper number 84-1323, 1984.
- Phillips, P. G.; Redd, B. "Propulsion Options for Manned Missions to the Moon and Mars", in Aerospace Century XXI (see Morgenthaler and Tobiska, under Propulsion above).
- Matloff, G. L. "Electric Propulsion and Interstellar Flight", 19th International Electric Propulsion Conference, Colorado Springs, Colorado, 11 May 1987.
- Korobeinikov, V. P. "On the Use of Solar Energy for the Acceleration of Bodies to Cosmic Velocities", Acta Astronautica, v 15 no 11 p 937-40, November 1987.
- Kerrebrock, J. L "Report of the National Commission on Space- One Commissioner's View", in Aerospace Century XXI (see Morgenthaler and Tobiska, under Propulsion above).
- Harvego, E. A.; Sulmeisters, T K. "A Comparison of Propulsion Systems for Potential Space Mission Applications", ASME Winter Meeting, Boston, Massachusetts, 13 December 1987, 1987.
- Byers, David C.; Wasel, Robert A. "NASA Electric Propulsion Program", NASA Technical Memorandum 89856, May 1987.

- Lorrey, M; **High Density Fuels**, web page, March 2006.
- Licht, S; **Solar Driven Synthesis**, in *Advanced Materials*, 2011.

Discussion Forums

Although online forums are informal, the collective knowledge of the participants is sometimes impressive. The discussion history can be searched or new questions asked to make use of it.

Appendix 1 - Fictional Methods

Fictional Methods Which Are Not Supported By Known Physics or Engineering

This section includes fictional methods which do not have any support from physics or engineering. Those that do have some supporting theory are listed in Part 2, Section 10 Theoretical Methods

Pre-Industrial Age

c. 1st Century

- Pseudo-Apollodorus, **Bibliotheca**, manuscript approximately 1st century. Contains the story of Daedalus and Icarus, who use wings made of feathers and wax. This fails on engineering due to the weakness of wax and power/weight ratio of humans. Feathers obviously work for birds, but they weigh much less than people do, and the feathers are attached directly to the skin, from which they grow.

1657

- de Bergerac, Cyrano, **A Voyage to the Moon**, 1657. Uses glass containers of dew for lift when the Sun falls on them in a first attempt. Dew is just condensed water, and exposing it to sunlight in nature simply evaporates it. In a closed container it would evaporate or simply sit there as a liquid, depending how much was there. In a later chapter ranks of firecrackers are used, which are not imaginary, but merely too low in energy.

18th and 19th Centuries

Cavorite

20th and 21st Centuries

Stutterwarp Drive

Fictional Methods Which Are Supported by Known Physics or Engineering

This section includes methods which have appeared in fictional works, but which are possible based on currently known science and technology.

Appendix 2 - Reference Data

Aerodynamics

- **US Air Force Stability and Control Data Compendium** - Starting in the 1920's, the US Air Force collected data for predicting the aerodynamic behavior of aircraft. This link is to the 1978 text version of the compiled data, known as the **USAF DATCOM**, which is a 3134 page, 113 MB .pdf file. Since then, versions of the data and formulas have been compiled into software modules, which can be run by themselves, or linked into more comprehensive airplane design software.

- **US Standard Atmosphere 1976** - This book gives information and tables describing composition, pressure, temperature, and other properties of the Earth's atmosphere under "standard" conditions. The actual atmosphere varies from these standard conditions due to weather, the solar cycle, and secular changes such as increasing CO_2 concentration. The reference data is still useful for doing calculations, and formulas based on this data have been incorporated into software programs and modules within programs.

Nuclear Rockets

Previous US Nuclear Rocket Program

Solid Core NTR Development

McDonnell Douglas has a long association with nuclear propulsion, beginning with nuclear aircraft propulsion studies in the mid-1950s.[1]. Additional work includes NERVA nuclear vehicle interplanetary flight studies, and solid core vehicle integration studies under contract to NASA in the 1960s. In 1971 studies were conducted on a Shuttle-launched Nuclear Shuttle System, which used the PEEWEE reactor.

In 1972 LANL undertook a study of a small nuclear engine using NERVA technology and based on the PEEWEE reactor design, that could be used in the newly designed NASA Space Shuttle.[2] This engine would have weighed 2,555 Kg, using composite UC-ZrC-C fuel, with other parameters including the following:

- Total Operating Period hours 2 1
- Operating Cycles 20 3
- Specific Impulse seconds 860 875
- Thrust kN 71.6 73.0
- Hydrogen Flow kg/sec 8.5 8.5
- Thermal Power MWt 354 367

Various other NERVA-derived propulsion concepts were considered, and LANL evaluated a range of mission applications. For instance, it was concluded that a nuclear-propelled orbital transfer vehicle could significantly reduce propellant launch requirements, based on mission models that envisioned from 145 to 872 Orbital Transfer Vehicle flights. The authors concluded that developments in non-nuclear nozzle and turbo-pump technology in the Space Shuttle Main Engine program have enhanced the viability of NERVA derived engines.

LANL mission analyses have identified significant reductions in Initial Mass in Low Earth Orbit (IMLEO) through the use of NERVA technology.[3]

Propulsion System	IMLEO (tons)
Chemical (all propulsive)	2,100
Chemical + Aerobrake	715
NTR (all propulsive)	760
NTR + Earth Only Aerobrake	540
NTR + Aerobrake	420

The costs of developing a flight-ready nuclear rocket was estimated to be $4-5 billion, which included the cost of rebuilding the 1960s NERVA capability, which was estimated to include:

Component	Cost	Notes
Engine Design & Construction	$1,218 M	80% of NERVA
Technology	$ 377 M	50% of NERVA
Test Facilities Capital	$ 460 M	
Operations	$ 210 M	
TOTAL	$2,265 M	in 1985 $

This analysis also considered issues associated with testing, noting that:

"The major obstacle to testing at NTS will be the reduced levels of radioactive debris which are allowed to transport into the public domain. The levels are more stringent than those present during the NERVA program. The current exposure limits of 150 m Rem to civilian personnel may restrict the tests of the NTR to low power levels and mass flows in the reactor... A simple solution to this problem may be to utilize one of the Pacific Ocean Islands owned by the United States -- namely Johnston Island... (an) ecological desert of ocean surrounds the area due to the stagnation of the return of the Japanese current..."

ANL Cermet Program

From 1961 through 1967 Argonne National Laboratory conducted a development program for nuclear rocket fuel elements independently of the Rover/NERVA program.[4] Although no engines were fabricated or tested, extensive work was conducted on testing ceramic fuel configurations. Thermal shock tests suggested that these cermet fuels could have substantial tolerance to the effects of nuclear excursions. This program was terminated in 1967, prior to its completion.

Evaluation

Two major problems were identified during the Rover/NERVA program:

Core disassembly due to vibration, accompanied by cracking of the fuel matrix and loss of material into the propellant flow;

Loss of fuel matrix uranium and carbon due to coating erosion and cracking, and through diffusion through the coating.

The first problem was resolved by changed designs which reduced vibration and matrix cracking.

However, the second problem, of fuel element corrosion, proved less tractable.[4]

"Corrosion was most pronounced in the mid-range region, about a third of the distance from the cold end of the fuel element. Fuel operating temperatures were lower here than the fabrication temperatures, hence thermal stresses were higher than at the hot end. Also, the neutron flux was highest in this region..."

"No fuel element geometry or fuel material ever totally solved the NERVA fuel element degradation problem. Mass loss of both uranium and carbon continued to limit service life by causing significant perturbation to core neutronics during the tests. Crack development in the fuel element coating was never completely eliminated.... Non-nuclear testing of coated fuel elements revealed an Arrhenius relationship between diffusion and temperature. For every 205 K increase in temperature (in the range 2400 to 2700 K), the mass loss increased by a factor of ten... resulting in loss of 20% of total uranium in approximately 5 hours of testing at 2870 K."

A number of other program management lessons emerged from this period. One analysis notes that:[4]

"One overriding lesson from the NERVA program is that fuel and core development should not be tied simply to a series of engine tests which require expensive nuclear operation. Definitive techniques for fuel evaluation in loops or in non-nuclear heated devices should be developed early and used throughout the program..."

Solid Core NTR Tests

The initial series of nuclear rocket tests was conducted by Los Alamos National Laboratory, under the KIWI program, which eventually cost $177 million (in then year dollars).

Graphite was chosen as the internal structural material for these nuclear reactors for several reasons. It has excellent strength at high temperatures and its strength actually increases at higher temperatures. In addition, in contrast to metals which are strong neutron absorbers, graphite acts as a moderator, reducing the amount of enriched uranium required in the core. The great disadvantage of graphite, however, is that it quickly erodes in the presence of hot hydrogen. While this erosion cannot be eliminated, techniques are available to reduce erosion to acceptable levels over the operating lifetime of the reactor, which is measured in minutes to hours.

Westinghouse was the prime contractor for the reactor component of the NERVA program, while Aerojet was the contractor for engine elements such as pumps and nozzles. The NERVA program resulted in an investment of $662 million (then-year dollars) for development and testing of flight engine prototypes.[4]

The first nuclear rocket test, conducted in July 1959, used uncoated UO2 plates as fuel elements. This test reached a maximum temperature of 2683 K and a power level of 70 MWt. Vibrations during operations produced significant structural damage in the reactor core.

The first nuclear reactor tested, KIWI-A, successfully demonstrated the principle of nuclear rockets, but it used unclad fuel plates that were not representative of later tests.

KIWI A test
: This test, conducted in July 1959, incorporated significant modifications in the core design used in KIWI A. The fuel consisted of short cylindrical Uranium Oxide elements in graphite modules, with four axial channels coated with Niobium Carbide using chemical vapor deposition. The reactor ran for 6 minutes at power levels as high as 85 MWt.

KIWI-A prime
: tested in 1960, replaced the fuel plates with NbC plated graphite modules with 4 micron diameter UO2 particles embedded in the graphite matrix. However, some structural damage occurred in this improved design during its 6 minute test.

KIWI A3
: The subsequent KIWI-A3 reactor used a higher temperature Chemical Vapor Deposition process, resulting in a thicker NbC coat with improved adherence. Some core damage occurred during the 5 minute test in October 1959, which reached power levels of 100 MWt, with some fuel elements showing blistering and corrosion. Generally this reactor test was considered successful.

KIWI B1A

Using the same UO2 fuel as KIWI A3, the fuel element design was changed to a 7 channel configuration 66 cm long, with niobium carbide coatings. The December 1961 test, which was intended to reach 1100 MWt, only reached 300 MWt, and was terminated after 30 seconds due to a fire caused by a hydrogen leak in the reactor exhaust nozzle.

KIWI B1B

This September 1962 test, which was essentially a repeat of the KIWI B1A test, achieved a power level of 900 MWT, but was terminated within a few seconds when several fuel elements were ejected from the reactor exhaust nozzle.

KIWI B2

Design configuration not tested.

KIWI B3

Design configuration not tested.

KIWI B4A

This reactor incorporated substantial redesigns based on the failure of the KIWI B1B configuration. The fuel elements were fully extruded hexagonal graphite blocks 1.32 meters long and 19 millimetres in diameter, with 19 cooling channels, each 2.3 millimetres in diameter. However, the test run in November 1962 was terminated when bright flashes in the exhaust stream indicated vibration induced damage to the core was leading to its disintegration.

KIWI B4B

Design configuration not tested.

KIWI B4C

Design configuration not tested.

KIWI B4D

Although modifications in the design of this reactor eliminated the vibration which had marred previous tests, the May 1964 test was terminated after about 60 seconds at full power due to the rupture of a nozzle cooling tube.

KIWI B4E

This reactor was the first to use coated Uranium Carbide (UC2) fuel, in place of the Uranium Oxide (UO2) fuel previously used. To avoid oxidation of the Carbide fuel, it the uranium fuel particles were coated with a 25 micron layer of pyrolytic graphite which exuded water vapor. This pyrolytic carbon layer subsequently was also used to enhance fission product retention, though this was not initially its purpose. This reactor was operated for 12 minutes, including 8 at full power. The duration of the test was limited by the available liquid hydrogen storage capability.

KIWI TNT

This KIWI-B type reactor was deliberately destroyed on January 1965 by subjecting it to a fast excursion. This test was intended to confirm theoretical models of transient behaviour.

Phoebus 1A

This first test of a new class of reactors on June 1965 included over 10 minutes of operations at 1090 MWt, with an exhaust temperature of 2370 K.

Phoebus 1B

This February 1967 test built on the previous Phoebus 1A test, reaching power levels of 1500 MWt for 30 minutes, with an additional 15 minutes at lower power levels.

Phoebus 2A

The most powerful nuclear reactor of any type ever constructed, with a design power level of 5,000 MWt. Operations in June 1968 were limited to 4,000 MWt dur to premature overheating of of aluminium segments of pressure vessel clamps. At total of 12.5 minutes of operations at temperatures of up to 2310 K included intermediate power level operations and reactor restart.

PEWEE

This small reactor was intended to be a reactor testbed, incorporating Zirconium Carbide coatings on some fuel elements, instead of the Niobium Carbide used on Phoebus. With a peak operating power of 503 MWt at 2550 K, it achieved new levels of core power density (2340 MWt/m3 average and 5200 MWt/m3 peak), demonstrating a specific impulse of 845 seconds.

Nuclear Furnace 1

The final phase of NERVA fuel development was the Nuclear Furnace (NF-1) reactor, a heterogeneous water-moderated beryllium-reflected reactor for high-temperature nuclear testing of fuel elements and other components. 47 of the 49 fuel elements used Uranium Carbide and Zirconium Carbide Carbon composite fuel, while the remaining 2 fuel elements used Uranium-Zirconium Carbide. Tests in 1972 of composite fuel elements with various carbide contents, thermal expansion coefficients, and thermal stress resistance demonstrated that minimizing the mismatch in thermal expansion coefficient between the fuel and coating would reduce coating cracking and carbon erosion. With a peak operating power of 44 MWt at 2500 K, it achieved new levels of core power density (4500 to 5000 MWt/m3).

The Nuclear Furnace test facility included provisions for remote controlled replacement of core elements, as well as a reactor effluent scrubber system to remove radioactive contaminants from the propellant exhaust.

Nuclear Furnace 2

This reactor was built but not tested, due to the cancellation of all work in this area in 1972. The goals of this experiment included testing of a novel coated particle fuel, using a graphite fuel matrix with a coefficient of thermal expansion closely matched to that of the coating, to reduce thermal stress and cracking.

NRX-A1

NRX-A2

This September 1964 engine test included 5 minutes of operation at half to full power levels. Duration of the test was limited by available hydrogen storage capacity. At full power of 1100 MWt the engine demonstrated a specific impulse of 760 seconds.

NRX-A3

This April 1965 test operated for 8 minutes, including 3.5 minutes at full power. The first test was terminated by a spurious trip from the turbine overspeed circuit. The reactor was restarted on May 1965, and operated at full power for 13 minutes, and subsequently restarted for low to medium power operations for an additional 45 minutes. In total, 66 minutes of operations were accumulated, including 16.5 minutes at full power.

NRX-EST

This engine operated for a total of 110 minutes, including 28 minutes at full power of 1100-1200 MWt, on five different days in February 1966.

NRX-A5

This 1100 MWt engine was operated in June 1966 for 30 minutes at full power, with the test duration limited by available hydrogen storage capacity.

NRX-A6

This 1100 MWt engine was operated in December 1967 for 60 minutes at full power, exceeding the NERVA design goal.

XECF

XE'

This 1100 MWt engine was a prototype engine, the first to operated in a downward firing position. It accumulated a total of 28 start cycles in March 1968 for a total of 115 minutes of operations. Test stand coolant water storage capacity limited each full power test to about 10 minutes.

1. Haloulakos, V.E., et al. *Nuclear Propulsion: Past, Present, and Future, Fifth Symposium on Space Nuclear Power Systems, Albuquerque 11-14 January 1988* pp. 329-332.
2. Bohl, R.J., and Boudreau, J.E. (January 1987) *Direct Nuclear Propulsion: A White Paper*. Los Alamos National Laboratory.
3. Howe, Steven (10-14 June 1985). *Assessment of the Advantages and Feasibility of a Nuclear Rocket for a Manned Mars Mission, Manned Mars Mission Workshop*. Huntsville, AL: Marshall Space Flight Center. Preprint LA-UR-85-2442.
4. Horman, F.J., et al (4-6 September 1991). *Particle Fuels Technology for Nuclear Thermal Propulsion, AIAA/NASA/OAI Conference on Advanced SEI Technologies*. Cleveland, Ohio. Paper AIAA 91-3457.
